

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 September 2008 (04.09.2008)

PCT

(10) International Publication Number
WO 2008/104598 A2

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)

Seoul 151-921 (KR). **OH, Jonghee** [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu, Seoul 151-921 (KR).

(21) International Application Number:
PCT/EP2008/052450

(74) Agent: **MISTRY, Meeta**; Basf Se, Global Intellectual Property, Gvx - C006, 67056 Ludwigshafen (DE).

(22) International Filing Date:
28 February 2008 (28.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
07103271.8 28 February 2007 (28.02.2007) EP
07104197.4 15 March 2007 (15.03.2007) EP
60/896,050 21 March 2007 (21.03.2007) US
60/909,510 2 April 2007 (02.04.2007) US

(71) Applicants (for all designated States except US):
CROPDESIGN N.V. [BE/BE]; Technolopark 3, B-9052 Zwijnaarde (BE). **CROP FUNCTIONAL GENOMICS CENTER** [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu, Seoul 151-921 (KR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHOI, Yang Do** [KR/KR]; Banppo Tower Apt 102-301, Jamwon-dong, Seocho-ku, Seoul 137-909 (KR). **PARK, Youn-II** [KR/KR]; Jijok dong 919, 705-401, Youseong, Daejeon 305-330 (KR). **CHUNG, Suck Won** [KR/KR]; Noeun 559-10, Youseong, Daejeon 305-325 (KR). **HWANG, In Gyu** [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu,

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND A METHOD FOR MAKING THE SAME

(57) Abstract: The present invention relates generally to the field of molecular biology and concerns a method for enhancing various economically important yield-related traits in plants. More specifically, the present invention concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a Harpin-associated Factor G polypeptide (hereinafter termed HpaG"). The present invention also concerns plants having modulated expression of a nucleic acid encoding an HpaG polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs comprising HpaG-encoding nucleic acids, useful in performing the methods of the invention. The present invention also provides a method for enhancing yield-related traits in plants relative to control plants, by modulating (preferably increasing) expression in a plant of a nucleic acid sequence encoding a SWITCH 2/ SUCROSE NON-FERMENTING 2 (SWI2/SNF2) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs useful in performing the methods of the invention.

WO 2008/104598 A2

Plants having enhanced yield-related traits and a method for making the same

The present invention relates generally to the field of molecular biology and concerns a method for enhancing various economically important yield-related traits in plants. More specifically, the present invention concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a Harpin-associated Factor G polypeptide (hereinafter termed "HpaG"). The present invention also concerns plants having modulated expression of a nucleic acid encoding an HpaG polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs comprising HpaG-encoding nucleic acids, useful in performing the methods of the invention. The present invention also provides a method for enhancing yield-related traits in plants relative to control plants, by modulating (preferably increasing) expression in a plant of a nucleic acid sequence encoding a SWITCH 2/ SUCROSE NON-FERMENTING 2 (SWI2/SNF2) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs useful in performing the methods of the invention.

The ever-increasing world population and the dwindling supply of arable land available for agriculture fuels research towards increasing the efficiency of agriculture. Conventional means for crop and horticultural improvements utilise selective breeding techniques to identify plants having desirable characteristics. However, such selective breeding techniques have several drawbacks, namely that these techniques are typically labour intensive and result in plants that often contain heterogeneous genetic components that may not always result in the desirable trait being passed on from parent plants. Advances in molecular biology have allowed mankind to modify the germplasm of animals and plants. Genetic engineering of plants entails the isolation and manipulation of genetic material (typically in the form of DNA or RNA) and the subsequent introduction of that genetic material into a plant. Such technology has the capacity to deliver crops or plants having various improved economic, agronomic or horticultural traits.

A trait of particular economic interest is increased yield. Yield is normally defined as the measurable produce of economic value from a crop. This may be defined in terms of quantity and/or quality. Yield is directly dependent on several factors, for example, the number and size of the organs, plant architecture (for example, the number of branches), seed production, leaf senescence and more. Root development, nutrient uptake, stress tolerance and early vigour may also be important factors in determining yield. Optimizing the abovementioned factors may therefore contribute to increasing crop yield.

Seed yield is a particularly important trait, since the seeds of many plants are important for human and animal nutrition. Crops such as, corn, rice, wheat, canola and soybean account for over half the total human caloric intake, whether through direct consumption of the seeds themselves or through consumption of meat products raised on processed seeds. They are also a source of sugars, oils and many kinds of metabolites used in industrial processes. Seeds contain an embryo (the source of new shoots and roots) and an endosperm (the source of nutrients for embryo growth during germination and during early growth of seedlings). The development of a seed involves many genes, and requires the transfer of metabolites from the roots, leaves and stems into the growing seed. The endosperm, in particular, assimilates the metabolic precursors of carbohydrates, oils and proteins and synthesizes them into storage macromolecules to fill out the grain.

Harvest index, the ratio of seed yield to aboveground dry weight, is relatively stable under many environmental conditions and so a robust correlation between plant size and grain yield can often be obtained (e.g. Rebetzke *et al.* (2002) Crop Science 42:739). These processes are intrinsically linked because the majority of grain biomass is dependent on current or stored photosynthetic productivity by the leaves and stem of the plant (Gardener *et al.* (1985) Physiology of Crop Plants. Iowa State University Press, pp 68-73). Therefore, selecting for plant size, even at early stages of development, has been used as an indicator for future potential yield (e.g. Tittonell *et al.* (2005) Agric Ecosys & Environ 105: 213). When testing for the impact of genetic differences on stress tolerance, the ability to standardize soil properties, temperature, water and nutrient availability and light intensity is an intrinsic advantage of greenhouse or plant growth chamber environments compared to the field. However, artificial limitations on yield due to poor pollination due to the absence of wind or insects, or insufficient space for mature root or canopy growth, can restrict the use of these controlled environments for testing yield differences. Therefore, measurements of plant size in early development, under standardized conditions in a growth chamber or greenhouse, are standard practices to provide indication of potential genetic yield advantages.

Another trait of particular economic interest is that of enhanced yield-related traits of plants grown under abiotic stress conditions. Abiotic stress is a primary cause of crop loss worldwide, reducing average yields for most major crop plants by more than 50% (Wang *et al.*, Planta (2003) 218: 1-14). Abiotic stresses may be caused by drought, salinity, temperature extremes, chemical toxicity and oxidative stress. The ability to enhance yield-related traits in plants grown under abiotic stress conditions would be of great economic advantage to farmers

worldwide and would allow for the cultivation of crops during adverse conditions and in territories where cultivation of crops may not otherwise be possible.

The ability to increase plant yield would have many applications in areas such as agriculture, including in the production of ornamental plants, arboriculture, horticulture and forestry. Increasing yield may also find use in the production of algae for use in bioreactors (for the biotechnological production of substances such as pharmaceuticals, antibodies or vaccines, or for the bioconversion of organic waste) and other such areas.

Background

I. HARPIN

The Type III Secretion System (TTSS) is an exporting machinery specific for Gram-negative bacteria and is found among plant and animal pathogens, but also in endosymbiotic *Rhizobia*. TTSS is postulated to deliver proteins into the host cell to which the bacterium is associated.

In plant pathogenic bacteria, the TTSS is a cluster of hypersensitive response and pathogenicity genes comprising about 20 genes, the Hrp cluster. Nine of these genes (the harpin conserved or *hrc*) are conserved among both plant and animal pathogens, eight of them share homology with genes encoding the flagella apparatus (Bogdanove et al., Mol. Microbiol. 20, 681-683, 1996), the ninth, *hrcC*, is homologous to the GSP outer membrane secretins (Deng and Huang, J. Bacteriol. 180, 4523-4531, 1999). The *hpa* (*hrp*-associated) genes contribute to pathogenicity and to the induction of the hypersensitive response (HR) in nonhost plants, but are not essential for the pathogenic interactions of bacteria with plants. The flagella apparatus and the TTSS are postulated to be evolved from a common origin (Gophna et al., Gene 312, 151-163, 2003); the TTSS has furthermore spread among evolutionary distant bacterial species via multiple horizontal-transfer events (Nguyen et al., J. Mol. Microbiol. Biotechnol. 2, 125-144, 2000).

Many gram-negative plant-pathogenic bacteria possess two sets of genes that modulate their interactions with plants. The avirulence genes determine host specificity based on gene-for-gene interactions, and the *hrp* (hypersensitive reaction and pathogenicity) genes are involved in pathogenicity and the induction of hypersensitive responses (HR) in nonhost plants. The HR is a highly localized plant cell death that occurs when non-host plants or resistant cultivars of host plants are infiltrated with the plant pathogen or HR elicitor molecules, such as Avr proteins and harpins. The HR is thought to be a resistance reaction of plants to microbial pathogens.

Harpins are a group of HR elicitors that are secreted by the type III secretion pathway (TTSS) and elicit HR when infiltrated into the apoplast of leaves of non-host plants. Unlike Avr proteins, which must be delivered inside the cell to exert their functions, harpins can elicit HR when delivered to the intercellular space of plant cells. Since the first harpin, HrpN, was identified from *Erwinia amylovora*, many harpins have been reported from various species, including *Pseudomonas*, *Ralstonia*, and *Xanthomonas*. Harpins are glycine-rich, heat stable proteins, lacking cysteine, and are postulated to be present in all plant pathogenic bacteria having a TTSS (Alfano and Colmer, Annu. Rev. Phytopathol. 42, 385-414, 2004). The biochemical mechanism of HR elicitation by harpins in non-host plants remains unclear. HrpZ of *Pseudomonas syringae* pv. *syringae* associates with the cell walls rather than the membranes of plant cells, and the protein elicits no response from protoplasts, which lack walls (Hoyos et al. Mol. Plant-Microbe Interact. 9, 608-616, 1996). However, HrpZ of *P. syringae* pv. *phaseolicola* binds to lipid bilayers and forms an ion-conducting pore (Lee et al., Proc. Natl. Acad. Sci. USA 98, 289-294, 2001). The N-terminal 109 amino acids and the C-terminal 216 amino acids of HrpZ are able to elicit HR to a level similar to full-length HrpZ (Alfano et al., Mol. Microbiol. 19, 715-728, 1996). Kim et al. and Charkowski et al. showed that the HrpW harpins of *E. amylovora* and *P. syringae* pv. *tomato* are composed of two domains—the N-terminal harpin domain and C-terminal Pel (pectate lyase) domain—and proposed that HrpW acts in the cell wall (Charkowski et al., J. Bacteriol. 180, 5211-5217, 1998; Kim and Beer, J. Bacteriol. 180, 5203-5210, 1998).

Besides harpins, the TTSS cluster in bacteria may also include genes encoding Harpin associated Factors. HpaG polypeptides are smaller than harpins, and they share little sequence homology. These sequence differences with harpins are postulated to contribute to the difference in the ability to elicit HR in plants between HpaG polypeptides and harpins (Kim et al., J. Bacteriol. 186, 6239-6247, 2004)

Korean patent application KR20030068302 discloses the *Xanthomonas* HpaG protein, which, when applied to plants or plant seeds, confers disease resistance, in particular resistance to *Xanthomonas axonopodis* infection. Harpin associated Factors have been used to confer disease resistance in plants; and as a result of this biotic stress resistance, plants had better yield compared to the control plants under biotic stress conditions.

Surprisingly it has now been found that modulating expression in a plant of a nucleic acid encoding a Harpin-associated Factor G polypeptide (HpaG) give plants enhanced yield-related traits relative to control plant. These enhanced yield-related traits were obtained in plants that were not exposed to stress.

II. SNF2

The present invention concerns a method for enhancing yield-related traits in plants relative to control plants by increasing expression in a plant of a nucleic acid sequence encoding a SWITCH 2/ SUCROSE NON-FERMENTING 2 (SWI2/SNF2) polypeptide.

5

Many chromosome-associated cellular processes, such as replication, transcription, DNA repair, or recombination, require accessible DNA. To deal with these events, cells possess activities that can remodel chromatin in eukaryotes or disrupt other DNA:protein complexes in both pro- and eukaryotes, using ATP hydrolysis. One of the best-studied examples of these activities is carried out by the SWI2/SNF2 family of ATPases, a large group of proteins implicated in many different remodeling-like processes.

10

SWI2/SNF2 family proteins are ubiquitous, as they are found in bacteria, archaea and eukaryotes. They have recently been classified into 24 distinct subfamilies, after multiple sequence alignment of the SWI2/SNF2 ATPase domain comprising the seven conserved sequence motifs (I, Ia, II, III, IV, V, and VI) (Flaus *et al.* (2006) *Nucleic Acids Res.* 2006; 34(10): 2887–2905). These subfamilies have traditionally taken the name of the archetypal member. One subfamily is named SSO1653, after the sole SWI2/SNF2 family member in archaeal *Sulfolobus solfataricus* (Flaus *et al.*, *supra*; Duur *et al.* (2005) *Cell* 121(3): 363-373), the uniquely archaeal and eubacterial subfamily most similar to the eukaryotic SWI2/SNF2 proteins. The SSO1653 subfamily carries all the SWI2/SNF2 family sequence and structural hallmarks.

15

20

US patent application US2003/233670 describes polynucleotides and proteins encoded by the polynucleotides. SEQ ID NO: 125 is a polynucleotide sequence encoding a SWI2/SNF2 polypeptide of the SSO1653 subfamily from *Synechocystis* sp. PCC 6803. US patent application US2005/108791 describes 24149 nucleic acid and polypeptide sequences, among which a nucleic acid sequence represented by SEQ ID NO: 57 encoding a SWI2/SNF2 polypeptide of the SSO1653 subfamily from *Synechocystis* sp. PCC 6803, as represented by SEQ ID NO: 396.

25

30

Surprisingly, it has now been found that increasing expression in a plant of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide gives plants having enhanced yield-related traits relative to control plants.

35

Definitions

Polypeptide(s)/Protein(s)

The terms "polypeptide" and "protein" are used interchangeably herein and refer to amino acids in a polymeric form of any length.

5

Polynucleotide(s)/Nucleic acid(s)/Nucleic acid sequence(s)/nucleotide sequence(s)

The terms "polynucleotide(s)", "nucleic acid sequence(s)", "nucleotide sequence(s)" are used interchangeably herein and refer to nucleotides, either ribonucleotides or deoxyribonucleotides or a combination of both, in a polymeric form of any length.

10

Control plant(s)

The choice of suitable control plants is a routine part of an experimental setup and may include corresponding wild type plants or corresponding plants without the gene of interest. The control plant is typically of the same plant species or even of the same variety as the plant to be assessed. The control plant may also be a nullizygote of the plant to be assessed. A "control plant" as used herein refers not only to whole plants, but also to plant parts, including seeds and seed parts.

15

Homologue(s)

"Homologues" of a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived.

20

A deletion refers to removal of one or more amino acids from a protein.

25

An insertion refers to one or more amino acid residues being introduced into a predetermined site in a protein. Insertions may comprise N-terminal and/or C-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than N- or C-terminal fusions, of the order of about 1 to 10 residues. Examples of N- or C-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, (histidine)-6-tag, glutathione S-transferase-tag, protein A, maltose-binding protein, dihydrofolate reductase, Tag•100 epitope, c-myc epitope, FLAG®-epitope, lacZ, CMP (calmodulin-binding peptide), HA epitope, protein C epitope and VSV epitope.

30

35

A substitution refers to replacement of amino acids of the protein with other amino acids having similar properties (such as similar hydrophobicity, hydrophilicity, antigenicity, propensity to form or break α -helical structures or β -sheet structures). Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; insertions will usually be of the order of about 1 to 10 amino acid residues. The amino acid substitutions are preferably conservative amino acid substitutions. Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company and Table 1 below).

Table 1: Examples of conserved amino acid substitutions

Residue	Conservative Substitutions	Residue	Conservative Substitutions
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser; Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Amino acid substitutions, deletions and/or insertions may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulation. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those skilled in the art and include M13 mutagenesis, T7-Gen *in vitro* mutagenesis (USB, Cleveland, OH), QuickChange Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols.

Derivatives

“Derivatives” include peptides, oligopeptides, polypeptides which may, compared to the amino acid sequence of the naturally-occurring form of the protein, such as the one presented in SEQ ID NO: 2, comprise substitutions of amino acids with non-naturally occurring amino acid residues, or additions of non-naturally occurring amino acid residues. “Derivatives” of a protein also encompass peptides, oligopeptides, polypeptides which comprise naturally occurring

altered (glycosylated, acylated, prenylated, phosphorylated, myristoylated, sulphated etc.) or non-naturally altered amino acid residues compared to the amino acid sequence of a naturally-occurring form of the polypeptide. A derivative may also comprise one or more non-amino acid substituents or additions compared to the amino acid sequence from which it is derived, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence, such as a reporter molecule which is bound to facilitate its detection, and non-naturally occurring amino acid residues relative to the amino acid sequence of a naturally-occurring protein.

10 Orthologue(s)/Parologue(s)

Orthologues and paralogues encompass evolutionary concepts used to describe the ancestral relationships of genes. Paralogues are genes within the same species that have originated through duplication of an ancestral gene and orthologues are genes from different organisms that have originated through speciation.

15

Domain

The term "domain" refers to a set of amino acids conserved at specific positions along an alignment of sequences of evolutionarily related proteins. While amino acids at other positions can vary between homologues, amino acids that are highly conserved at specific positions indicate amino acids that are likely essential in the structure, stability or activity of a protein. Identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers to determine if any polypeptide in question belongs to a previously identified polypeptide family.

20

25 Motif/Consensus sequence/Signature

The term "motif" or "consensus sequence" or "signature" refers to a short conserved region in the sequence of evolutionarily related proteins. Motifs are frequently highly conserved parts of domains, but may also include only part of the domain, or be located outside of conserved domain (if all of the amino acids of the motif fall outside of a defined domain).

30

Hybridisation

The term "hybridisation" as defined herein is a process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridisation process can occur entirely in solution, i.e. both complementary nucleic acids are in solution. The hybridisation process can also occur with one of the complementary nucleic acids immobilised to a matrix such as magnetic beads, Sepharose beads or any other resin. The hybridisation process can furthermore occur with one of the complementary nucleic acids immobilised to a

35

solid support such as a nitro-cellulose or nylon membrane or immobilised by e.g. photolithography to, for example, a siliceous glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). In order to allow hybridisation to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids.

The term "stringency" refers to the conditions under which a hybridisation takes place. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration, ionic strength and hybridisation buffer composition. Generally, low stringency conditions are selected to be about 30°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Medium stringency conditions are when the temperature is 20°C below T_m , and high stringency conditions are when the temperature is 10°C below T_m . High stringency hybridisation conditions are typically used for isolating hybridising sequences that have high sequence similarity to the target nucleic acid sequence. However, nucleic acids may deviate in sequence and still encode a substantially identical polypeptide, due to the degeneracy of the genetic code. Therefore medium stringency hybridisation conditions may sometimes be needed to identify such nucleic acid molecules.

The T_m is the temperature under defined ionic strength and pH, at which 50% of the target sequence hybridises to a perfectly matched probe. The T_m is dependent upon the solution conditions and the base composition and length of the probe. For example, longer sequences hybridise specifically at higher temperatures. The maximum rate of hybridisation is obtained from about 16°C up to 32°C below T_m . The presence of monovalent cations in the hybridisation solution reduce the electrostatic repulsion between the two nucleic acid strands thereby promoting hybrid formation; this effect is visible for sodium concentrations of up to 0.4M (for higher concentrations, this effect may be ignored). Formamide reduces the melting temperature of DNA-DNA and DNA-RNA duplexes with 0.6 to 0.7°C for each percent formamide, and addition of 50% formamide allows hybridisation to be performed at 30 to 45°C, though the rate of hybridisation will be lowered. Base pair mismatches reduce the hybridisation rate and the thermal stability of the duplexes. On average and for large probes, the T_m decreases about 1°C per % base mismatch. The T_m may be calculated using the following equations, depending on the types of hybrids:

- 1) DNA-DNA hybrids (Meinkoth and Wahl, Anal. Biochem., 138: 267-284, 1984):

$$T_m = 81.5^\circ\text{C} + 16.6 \times \log_{10}[\text{Na}^+]^a + 0.41 \times \%[\text{G/C}^b] - 500 \times [\text{L}^c]^{-1} - 0.61 \times \% \text{ formamide}$$
- 2) DNA-RNA or RNA-RNA hybrids:

$$T_m = 79.8 + 18.5 (\log_{10}[\text{Na}^+]^a) + 0.58 (\%G/C^b) + 11.8 (\%G/C^b)^2 - 820/L^c$$

3) oligo-DNA or oligo-RNA^d hybrids:

For <20 nucleotides: $T_m = 2 (\ln)$

For 20–35 nucleotides: $T_m = 22 + 1.46 (\ln)$

- 5 ^a or for other monovalent cation, but only accurate in the 0.01–0.4 M range.
 ^b only accurate for %GC in the 30% to 75% range.
 ^c L = length of duplex in base pairs.
 ^d Oligo, oligonucleotide; \ln , effective length of primer = $2 \times (\text{no. of G/C}) + (\text{no. of A/T})$.

- 10 Non-specific binding may be controlled using any one of a number of known techniques such as, for example, blocking the membrane with protein containing solutions, additions of heterologous RNA, DNA, and SDS to the hybridisation buffer, and treatment with Rnase. For non-homologous probes, a series of hybridizations may be performed by varying one of (i) progressively lowering the annealing temperature (for example from 68°C to 42°C) or (ii)
 15 progressively lowering the formamide concentration (for example from 50% to 0%). The skilled artisan is aware of various parameters which may be altered during hybridisation and which will either maintain or change the stringency conditions.

- Besides the hybridisation conditions, specificity of hybridisation typically also depends on the
 20 function of post-hybridisation washes. To remove background resulting from non-specific hybridisation, samples are washed with dilute salt solutions. Critical factors of such washes include the ionic strength and temperature of the final wash solution: the lower the salt concentration and the higher the wash temperature, the higher the stringency of the wash. Wash conditions are typically performed at or below hybridisation stringency. A positive
 25 hybridisation gives a signal that is at least twice of that of the background. Generally, suitable stringent conditions for nucleic acid hybridisation assays or gene amplification detection procedures are as set forth above. More or less stringent conditions may also be selected. The skilled artisan is aware of various parameters which may be altered during washing and which will either maintain or change the stringency conditions.

- 30 For example, typical high stringency hybridisation conditions for DNA hybrids longer than 50 nucleotides encompass hybridisation at 65°C in 1x SSC or at 42°C in 1x SSC and 50% formamide, followed by washing at 65°C in 0.3x SSC. Examples of medium stringency hybridisation conditions for DNA hybrids longer than 50 nucleotides encompass hybridisation
 35 at 50°C in 4x SSC or at 40°C in 6x SSC and 50% formamide, followed by washing at 50°C in 2x SSC. The length of the hybrid is the anticipated length for the hybridising nucleic acid. When nucleic acids of known sequence are hybridised, the hybrid length may be determined

by aligning the sequences and identifying the conserved regions described herein. 1×SSC is 0.15M NaCl and 15mM sodium citrate; the hybridisation solution and wash solutions may additionally include 5 × Denhardt's reagent, 0.5-1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.5% sodium pyrophosphate.

5

For the purposes of defining the level of stringency, reference can be made to Sambrook et al. (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York or to Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989 and yearly updates).

10

Gene shuffling/Directed evolution

Gene shuffling or directed evolution consists of iterations of DNA shuffling followed by appropriate screening and/or selection to generate variants of nucleic acids or portions thereof encoding proteins having a modified biological activity (Castle et al., (2004) Science 304(5674): 1151-4; US patents 5,811,238 and 6,395,547).

15

Regulatory element/Control sequence/Promoter

The terms “regulatory element”, “control sequence” and “promoter” are all used interchangeably herein and are to be taken in a broad context to refer to regulatory nucleic acid sequences capable of effecting expression of the sequences to which they are ligated. The term “promoter” typically refers to a nucleic acid control sequence located upstream from the transcriptional start of a gene and which is involved in recognising and binding of RNA polymerase and other proteins, thereby directing transcription of an operably linked nucleic acid. Encompassed by the aforementioned terms are transcriptional regulatory sequences derived from a classical eukaryotic genomic gene (including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence) and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. Also included within the term is a transcriptional regulatory sequence of a classical prokaryotic gene, in which case it may include a –35 box sequence and/or –10 box transcriptional regulatory sequences. The term “regulatory element” also encompasses a synthetic fusion molecule or derivative that confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ.

20

25

30

35

A “plant promoter” comprises regulatory elements, which mediate the expression of a coding sequence segment in plant cells. Accordingly, a plant promoter need not be of plant origin, but may originate from viruses or micro-organisms, for example from viruses which attack plant

cells. The “plant promoter” can also originate from a plant cell, e.g. from the plant which is transformed with the nucleic acid sequence to be expressed in the inventive process and described herein. This also applies to other “plant” regulatory signals, such as “plant” terminators. The promoters upstream of the nucleotide sequences useful in the methods of the present invention can be modified by one or more nucleotide substitution(s), insertion(s) and/or deletion(s) without interfering with the functionality or activity of either the promoters, the open reading frame (ORF) or the 3'-regulatory region such as terminators or other 3' regulatory regions which are located away from the ORF. It is furthermore possible that the activity of the promoters is increased by modification of their sequence, or that they are replaced completely by more active promoters, even promoters from heterologous organisms. For expression in plants, the nucleic acid molecule must, as described above, be linked operably to or comprise a suitable promoter which expresses the gene at the right point in time and with the required spatial expression pattern.

Operably linked

The term “operably linked” as used herein refers to a functional linkage between the promoter sequence and the gene of interest, such that the promoter sequence is able to initiate transcription of the gene of interest.

Constitutive promoter

A “constitutive promoter” refers to a promoter that is transcriptionally active during most, but not necessarily all, phases of growth and development and under most environmental conditions, in at least one cell, tissue or organ. Table 2a below gives examples of constitutive promoters.

Table 2a: Examples of constitutive promoters

Gene Source	Reference
Actin	McElroy et al, Plant Cell, 2: 163-171, 1990
HMGP	WO 2004/070039
CAMV 35S	Odell et al, Nature, 313: 810-812, 1985
CaMV 19S	Nilsson et al., Physiol. Plant. 100:456-462, 1997
GOS2	de Pater et al, Plant J Nov;2(6):837-44, 1992, WO 2004/065596
Ubiquitin	Christensen et al, Plant Mol. Biol. 18: 675-689, 1992
Rice cyclophilin	Buchholz et al, Plant Mol Biol. 25(5): 837-43, 1994
Maize H3 histone	Lepetit et al, Mol. Gen. Genet. 231:276-285, 1992
Alfalfa H3 histone	Wu et al. Plant Mol. Biol. 11:641-649, 1988
Actin 2	An et al, Plant J. 10(1); 107-121, 1996

34S FMV	Sanger et al., Plant. Mol. Biol., 14, 1990: 433-443
Rubisco small subunit	US 4,962,028
OCS	Leisner (1988) Proc Natl Acad Sci USA 85(5): 2553
SAD1	Jain et al., Crop Science, 39 (6), 1999: 1696
SAD2	Jain et al., Crop Science, 39 (6), 1999: 1696
Nos	Shaw et al. (1984) Nucleic Acids Res. 12(20):7831-7846
V-ATPase	WO 01/14572
Super promoter	WO 95/14098
G-box proteins	WO 94/12015

Ubiquitous promoter

A ubiquitous promoter is active in substantially all tissues or cells of an organism.

5 Developmentally-regulated promoter

A developmentally-regulated promoter is active during certain developmental stages or in parts of the plant that undergo developmental changes.

Inducible promoter

- 10 An inducible promoter has induced or increased transcription initiation in response to a chemical (for a review see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89-108), environmental or physical stimulus, or may be “stress-inducible”, i.e. activated when a plant is exposed to various stress conditions, or a “pathogen-inducible” i.e. activated when a plant is exposed to exposure to various pathogens.

15

Organ-specific/Tissue-specific promoter

- An organ-specific or tissue-specific promoter is one that is capable of preferentially initiating transcription in certain organs or tissues, such as the leaves, roots, seed tissue etc. For example, a “root-specific promoter” is a promoter that is transcriptionally active predominantly in plant roots, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts. Promoters able to initiate transcription in certain cells only are referred to herein as “cell-specific”.
- 20

Examples of root-specific promoters are listed in Table 2b below:

25

Table 2b: Examples of root-specific promoters

Gene Source	Reference
RCc3	Plant Mol Biol. 1995 Jan;27(2):237-48
Arabidopsis PHT1	Kovama et al., 2005; Mudge et al. (2002, Plant J. 31:341)
Medicago phosphate transporter	Xiao et al., 2006
Arabidopsis Pyk10	Nitz et al. (2001) Plant Sci 161(2): 337-346
root-expressible genes	Tingey et al., EMBO J. 6: 1, 1987.
tobacco auxin-inducible gene	Van der Zaal et al., Plant Mol. Biol. 16, 983, 1991.
β -tubulin	Oppenheimer, et al., Gene 63: 87, 1988.
tobacco root-specific genes	Conkling, et al., Plant Physiol. 93: 1203, 1990.
B. napus G1-3b gene	United States Patent No. 5, 401, 836
SbPRP1	Suzuki et al., Plant Mol. Biol. 21: 109-119, 1993.
LRX1	Baumberger et al. 2001, Genes & Dev. 15:1128
BTG-26 Brassica napus	US 20050044585
LeAMT1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
The LeNRT1-1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
class I patatin gene (potato)	Liu et al., Plant Mol. Biol. 153:386-395, 1991.
KDC1 (Daucus carota)	Downey et al. (2000, J. Biol. Chem. 275:39420)
TobRB7 gene	W Song (1997) PhD Thesis, North Carolina State University, Raleigh, NC USA
OsRAB5a (rice)	Wang et al. 2002, Plant Sci. 163:273
ALF5 (Arabidopsis)	Diener et al. (2001, Plant Cell 13:1625)
NRT2;1Np (N. plumbaginifolia)	Quesada et al. (1997, Plant Mol. Biol. 34:265)

A seed-specific promoter is transcriptionally active predominantly in seed tissue, but not necessarily exclusively in seed tissue (in cases of leaky expression). The seed-specific promoter may be active during seed development and/or during germination. The seed specific promoter may be endosperm and/or aleurone and/or embryo specific. Examples of seed-specific promoters (endosperm/aleurone/embryo specific) are shown in Table 2c, d, e, f below. Further examples of seed-specific promoters are given in Qing Qu and Takaiwa (Plant Biotechnol. J. 2, 113-125, 2004), which disclosure is incorporated by reference herein as if fully set forth.

Table 2c: Examples of seed-specific promoters

Gene source	Reference
seed-specific genes	Simon et al., Plant Mol. Biol. 5: 191, 1985;
	Scofield et al., J. Biol. Chem. 262: 12202, 1987.;
	Baszczynski et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	Pearson et al., Plant Mol. Biol. 18: 235-245, 1992.
Legumin	Ellis et al., Plant Mol. Biol. 10: 203-214, 1988.
glutelin (rice)	Takaiwa et al., Mol. Gen. Genet. 208: 15-22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43-47, 1987.
Zein	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	Stalberg et al, Planta 199: 515-519, 1996.
wheat LMW and HMW glutenin-1	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	Albani et al, Plant Cell, 9: 171-184, 1997
wheat α , β , γ -gliadins	EMBO J. 3:1409-15, 1984
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice a-globulin Glb-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
rice α -globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophos- phorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum α -kafirin	DeRose et al., Plant Mol. Biol 32:1029-35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Wu et al, J. Biochem. 123:386, 1998
sunflower oleosin	Cummins et al., Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	WO 2004/070039
PRO0136, rice alanine aminotransferase	unpublished

PRO0147, trypsin inhibitor ITR1 (barley)	unpublished
PRO0151, rice WSI18	WO 2004/070039
PRO0175, rice RAB21	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149:1125-38, 1998

Table 2d: examples of endosperm-specific promoters

Gene source	Reference
glutelin (rice)	Takaiwa et al. (1986) Mol Gen Genet 208:15-22; Takaiwa et al. (1987) FEBS Letts. 221:43-47
Zein	Matzke et al., (1990) Plant Mol Biol 14(3): 323-32
wheat LMW and HMW glutenin-1	Colot et al. (1989) Mol Gen Genet 216:81-90, Anderson et al. (1989) NAR 17:461-2
wheat SPA	Albani et al. (1997) Plant Cell 9:171-184
wheat gliadins	Rafalski et al. (1984) EMBO 3:1409-15
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Cho et al. (1999) Theor Appl Genet 98:1253-62; Muller et al. (1993) Plant J 4:343-55; Sorenson et al. (1996) Mol Gen Genet 250:750-60
barley DOF	Mena et al, (1998) Plant J 116(1): 53-62
blz2	Onate et al. (1999) J Biol Chem 274(14):9175-82
Synthetic promoter	Vicente-Carbajosa et al. (1998) Plant J 13:629-640
rice prolamin NRP33	Wu et al, (1998) Plant Cell Physiol 39(8) 885-889
rice globulin Glb-1	Wu et al. (1998) Plant Cell Physiol 39(8) 885-889
rice globulin REB/OHP-1	Nakase et al. (1997) Plant Molec Biol 33: 513-522
rice ADP-glucose pyrophosphorylase	Russell et al. (1997) Trans Res 6:157-68
maize ESR gene family	Opsahl-Ferstad et al. (1997) Plant J 12:235-46
Sorghum kafirin	DeRose et al. (1996) Plant Mol Biol 32:1029-35

Table 2e: Examples of embryo specific promoters:

Gene source	Reference
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
PRO0151	WO 2004/070039
PRO0175	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039

Table 2f: Examples of aleurone-specific promoters:

Gene source	Reference
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
Cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149:1125-38,1998

- 5 A green tissue-specific promoter as defined herein is a promoter that is transcriptionally active predominantly in green tissue, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts.

10 Examples of green tissue-specific promoters which may be used to perform the methods of the invention are shown in Table 2g below.

Table 2g: Examples of green tissue-specific promoters

Gene	Expression	Reference
Maize Orthophosphate dikinase	Leaf specific	Fukavama et al., 2001
Maize Phosphoenolpyruvate carboxylase	Leaf specific	Kausch et al., 2001
Rice Phosphoenolpyruvate carboxylase	Leaf specific	Liu et al., 2003
Rice small subunit Rubisco	Leaf specific	Nomura et al., 2000
rice beta expansin EXBP9	Shoot specific	WO 2004/070039
Pigeonpea small subunit Rubisco	Leaf specific	Panguluri et al., 2005
Pea RBCS3A	Leaf specific	

15 Another example of a tissue-specific promoter is a meristem-specific promoter, which is transcriptionally active predominantly in meristematic tissue, substantially to the exclusion of

any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts. Examples of green meristem-specific promoters which may be used to perform the methods of the invention are shown in Table 2h below.

5 **Table 2h:** Examples of meristem-specific promoters

Gene source	Expression pattern	Reference
rice OSH1	Shoot apical meristem, from embryo globular stage to seedling stage	Sato <i>et al.</i> (1996) Proc. Natl. Acad. Sci. USA, 93: 8117-8122
Rice metallothionein	Meristem specific	BAD87835.1
WAK1 & WAK 2	Shoot and root apical meristems, and in expanding leaves and sepals	Wagner & Kohorn (2001) Plant Cell 13(2): 303–318

Terminator

The term “terminator” encompasses a control sequence which is a DNA sequence at the end of a transcriptional unit which signals 3' processing and polyadenylation of a primary transcript and termination of transcription. The terminator can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The terminator to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

Selectable marker (gene)/Reporter gene

“Selectable marker”, “selectable marker gene” or “reporter gene” includes any gene that confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells that are transfected or transformed with a nucleic acid construct of the invention. These marker genes enable the identification of a successful transfer of the nucleic acid molecules via a series of different principles. Suitable markers may be selected from markers that confer antibiotic or herbicide resistance, that introduce a new metabolic trait or that allow visual selection. Examples of selectable marker genes include genes conferring resistance to antibiotics (such as nptII that phosphorylates neomycin and kanamycin, or hpt, phosphorylating hygromycin, or genes conferring resistance to, for example, bleomycin, streptomycin, tetracyclin, chloramphenicol, ampicillin, gentamycin, geneticin (G418), spectinomycin or blasticidin), to herbicides (for example bar which provides resistance to Basta®; aroA or gox providing resistance against glyphosate, or the genes conferring resistance to, for example, imidazolinone, phosphinothricin or sulfonylurea), or genes that provide a metabolic trait (such as manA that allows plants to use mannose as sole carbon

source or xylose isomerase for the utilisation of xylose, or antinutritive markers such as the resistance to 2-deoxyglucose). Expression of visual marker genes results in the formation of colour (for example β -glucuronidase, GUS or β -galactosidase with its coloured substrates, for example X-Gal), luminescence (such as the luciferin/luciferase system) or fluorescence (Green Fluorescent Protein, GFP, and derivatives thereof). This list represents only a small number of possible markers. The skilled worker is familiar with such markers. Different markers are preferred, depending on the organism and the selection method.

Transgenic/Transgene/Recombinant

For the purposes of the invention, "transgenic", "transgene" or "recombinant" means with regard to, for example, a nucleic acid sequence, an expression cassette, gene construct or a vector comprising the nucleic acid sequence or an organism transformed with the nucleic acid sequences, expression cassettes or vectors according to the invention, all those constructions brought about by recombinant methods in which either

- (a) the nucleic acid sequences encoding proteins useful in the methods of the invention, or
- (b) genetic control sequence(s) which is operably linked with the nucleic acid sequence according to the invention, for example a promoter, or
- (c) a) and b)

are not located in their natural genetic environment or have been modified by recombinant methods, it being possible for the modification to take the form of, for example, a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. The natural genetic environment is understood as meaning the natural genomic or chromosomal locus in the original plant or the presence in a genomic library. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained, at least in part.

The environment flanks the nucleic acid sequence at least on one side and has a sequence length of at least 50 bp, preferably at least 500 bp, especially preferably at least 1000 bp, most preferably at least 5000 bp. A naturally occurring expression cassette – for example the naturally occurring combination of the natural promoter of the nucleic acid sequences with the corresponding nucleic acid sequence encoding a polypeptide useful in the methods of the present invention, as defined above – becomes a transgenic expression cassette when this expression cassette is modified by non-natural, synthetic ("artificial") methods such as, for example, mutagenic treatment. Suitable methods are described, for example, in US 5,565,350 or WO 00/15815.

A transgenic plant for the purposes of the invention is thus understood as meaning, as above, that the nucleic acids used in the method of the invention are not at their natural locus in the genome of said plant, it being possible for the nucleic acids to be expressed homologously or

heterologously. However, as mentioned, transgenic also means that, while the nucleic acids according to the invention or used in the inventive method are at their natural position in the genome of a plant, the sequence has been modified with regard to the natural sequence, and/or that the regulatory sequences of the natural sequences have been modified.

5 Transgenic is preferably understood as meaning the expression of the nucleic acids according to the invention at an unnatural locus in the genome, i.e. homologous or, preferably, heterologous expression of the nucleic acids takes place. Preferred transgenic plants are mentioned herein.

10 Transformation

The term "introduction" or "transformation" as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated there from. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell may then be used to regenerate a transformed plant in a manner known to persons skilled in the art.

25 The transfer of foreign genes into the genome of a plant is called transformation. Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. et al., (1982) Nature 296, 72-74; Negrutiu I et al. (1987) Plant Mol Biol 8: 363-373); electroporation of protoplasts (Shillito R.D. et al. (1985) Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A et al., (1986) Mol. Gen Genet 202: 179-185); DNA or RNA-coated particle bombardment (Klein

TM et al., (1987) Nature 327: 70) infection with (non-integrative) viruses and the like. Transgenic plants, including transgenic crop plants, are preferably produced via Agrobacterium-mediated transformation. An advantageous transformation method is the transformation in planta. To this end, it is possible, for example, to allow the agrobacteria to act
5 on plant seeds or to inoculate the plant meristem with agrobacteria. It has proved particularly expedient in accordance with the invention to allow a suspension of transformed agrobacteria to act on the intact plant or at least on the flower primordia. The plant is subsequently grown on until the seeds of the treated plant are obtained (Clough and Bent, Plant J. (1998) 16, 735–743). Methods for Agrobacterium-mediated transformation of rice include well known methods
10 for rice transformation, such as those described in any of the following: European patent application EP 1198985 A1, Aldemita and Hodges (Planta 199: 612-617, 1996); Chan et al. (Plant Mol Biol 22 (3): 491-506, 1993), Hiei et al. (Plant J 6 (2): 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida et al. (Nat. Biotechnol
15 14(6): 745-50, 1996) or Frame et al. (Plant Physiol 129(1): 13-22, 2002), which disclosures are incorporated by reference herein as if fully set forth. Said methods are further described by way of example in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205-225). The nucleic acids
20 or the construct to be expressed is preferably cloned into a vector, which is suitable for transforming Agrobacterium tumefaciens, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711). Agrobacteria transformed by such a vector can then be used in known manner for the transformation of plants, such as plants used as a model, like Arabidopsis (Arabidopsis thaliana is within the scope of the present invention not considered as a crop
25 plant), or crop plants such as, by way of example, tobacco plants, for example by immersing bruised leaves or chopped leaves in an agrobacterial solution and then culturing them in suitable media. The transformation of plants by means of Agrobacterium tumefaciens is described, for example, by Höfgen and Willmitzer in Nucl. Acid Res. (1988) 16, 9877 or is known inter alia from F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic
30 Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

In addition to the transformation of somatic cells, which then have to be regenerated into intact plants, it is also possible to transform the cells of plant meristems and in particular those cells
35 which develop into gametes. In this case, the transformed gametes follow the natural plant development, giving rise to transgenic plants. Thus, for example, seeds of Arabidopsis are treated with agrobacteria and seeds are obtained from the developing plants of which a certain

proportion is transformed and thus transgenic [Feldman, KA and Marks MD (1987). Mol Gen Genet 208:274-289; Feldmann K (1992). In: C Koncz, N-H Chua and J Shell, eds, Methods in Arabidopsis Research. Word Scientific, Singapore, pp. 274-289]. Alternative methods are based on the repeated removal of the inflorescences and incubation of the excision site in the center of the rosette with transformed agrobacteria, whereby transformed seeds can likewise be obtained at a later point in time (Chang (1994). Plant J. 5: 551-558; Katavic (1994). Mol Gen Genet, 245: 363-370). However, an especially effective method is the vacuum infiltration method with its modifications such as the "floral dip" method. In the case of vacuum infiltration of Arabidopsis, intact plants under reduced pressure are treated with an agrobacterial suspension [Bechthold, N (1993). C R Acad Sci Paris Life Sci, 316: 1194-1199], while in the case of the "floral dip" method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension [Clough, SJ und Bent, AF (1998). The Plant J. 16, 735-743]. A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In addition the stable transformation of plastids is of advantages because plastids are inherited maternally in most crops reducing or eliminating the risk of transgene flow through pollen. The transformation of the chloroplast genome is generally achieved by a process which has been schematically displayed in Klaus et al., 2004 [Nature Biotechnology 22 (2), 225-229]. Briefly the sequences to be transformed are cloned together with a selectable marker gene between flanking sequences homologous to the chloroplast genome. These homologous flanking sequences direct site specific integration into the plastome. Plastidal transformation has been described for many different plant species and an overview is given in Bock (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol. 2001 Sep 21; 312 (3):425-38 or Maliga, P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol. 21, 20-28. Further biotechnological progress has recently been reported in form of marker free plastid transformants, which can be produced by a transient co-integrated marker gene (Klaus et al., 2004, Nature Biotechnology 22(2), 225-229).

TILLING

TILLING (Targeted Induced Local Lesions In Genomes) is a mutagenesis technology useful to generate and/or identify nucleic acids encoding proteins with modified expression and/or activity. TILLING also allows selection of plants carrying such mutant variants. These mutant variants may exhibit modified expression, either in strength or in location or in timing (if the mutations affect the promoter for example). These mutant variants may exhibit higher activity than that exhibited by the gene in its natural form. TILLING combines high-density mutagenesis with high-throughput screening methods. The steps typically followed in TILLING

are: (a) EMS mutagenesis (Redei GP and Koncz C (1992) In Methods in Arabidopsis Research, Koncz C, Chua NH, Schell J, eds. Singapore, World Scientific Publishing Co, pp. 16–82; Feldmann et al., (1994) In Meyerowitz EM, Somerville CR, eds, Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 137-172; Lightner J and Caspar T (1998) In J Martinez-Zapater, J Salinas, eds, Methods on Molecular Biology, Vol. 82. Humana Press, Totowa, NJ, pp 91-104); (b) DNA preparation and pooling of individuals; (c) PCR amplification of a region of interest; (d) denaturation and annealing to allow formation of heteroduplexes; (e) DHPLC, where the presence of a heteroduplex in a pool is detected as an extra peak in the chromatogram; (f) identification of the mutant individual; and (g) sequencing of the mutant PCR product. Methods for TILLING are well known in the art (McCallum et al., (2000) Nat Biotechnol 18: 455-457; reviewed by Stemple (2004) Nat Rev Genet 5(2): 145-50).

Yield

The term “yield” in general means a measurable produce of economic value, typically related to a specified crop, to an area, and to a period of time. Individual plant parts directly contribute to yield based on their number, size and/or weight, or the actual yield is the yield per acre for a crop and year, which is determined by dividing total production (includes both harvested and appraised production) by planted acres.

Increase/Improve/Enhance

The terms “increase”, “improve” or “enhance” are interchangeable and shall mean in the sense of the application at least a 5%, 6%, 7%, 8%, 9% or 10%, preferably at least 15% or 20%, more preferably 25%, 30%, 35% or 40% more yield and/or growth in comparison to control plants as defined herein.

Seed yield

Increased seed yield may manifest itself as one or more of the following: a) an increase in seed biomass (total seed weight) which may be on an individual seed basis and/or per plant and/or per hectare or acre; b) increased number of flowers per plant; c) increased number of (filled) seeds; d) increased seed filling rate (which is expressed as the ratio between the number of filled seeds divided by the total number of seeds); e) increased harvest index, which is expressed as a ratio of the yield of harvestable parts, such as seeds, divided by the total biomass; and f) increased thousand kernel weight (TKW), which is extrapolated from the number of filled seeds counted and their total weight. An increased TKW may result from an increased seed size and/or seed weight, and may also result from an increase in embryo and/or endosperm size.

An increase in seed yield may also be manifested as an increase in seed size and/or seed volume. Furthermore, an increase in seed yield may also manifest itself as an increase in seed area and/or seed length and/or seed width and/or seed perimeter. Increased yield may also result in modified architecture, or may occur because of modified architecture.

5

Plant

The term "plant" as used herein encompasses whole plants, ancestors and progeny of the plants and plant parts, including seeds, shoots, stems, leaves, roots (including tubers), flowers, and tissues and organs, wherein each of the aforementioned comprise the gene/nucleic acid of interest. The term "plant" also encompasses plant cells, suspension cultures, callus tissue, embryos, meristematic regions, gametophytes, sporophytes, pollen and microspores, again wherein each of the aforementioned comprises the gene/nucleic acid of interest.

Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs selected from the list comprising *Acer* spp., *Actinidia* spp., *Abelmoschus* spp., *Agave sisalana*, *Agropyron* spp., *Agrostis stolonifera*, *Allium* spp., *Amaranthus* spp., *Ammophila arenaria*, *Ananas comosus*, *Annona* spp., *Apium graveolens*, *Arachis* spp., *Artocarpus* spp., *Asparagus officinalis*, *Avena* spp. (e.g. *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida*), *Averrhoa carambola*, *Bambusa* sp., *Benincasa hispida*, *Bertholletia excelsea*, *Beta vulgaris*, *Brassica* spp. (e.g. *Brassica napus*, *Brassica rapa* ssp. [canola, oilseed rape, turnip rape]), *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Cannabis sativa*, *Capsicum* spp., *Carex elata*, *Carica papaya*, *Carissa macrocarpa*, *Carya* spp., *Carthamus tinctorius*, *Castanea* spp., *Ceiba pentandra*, *Cichorium endivia*, *Cinnamomum* spp., *Citrullus lanatus*, *Citrus* spp., *Cocos* spp., *Coffea* spp., *Colocasia esculenta*, *Cola* spp., *Corchorus* sp., *Coriandrum sativum*, *Corylus* spp., *Crataegus* spp., *Crocus sativus*, *Cucurbita* spp., *Cucumis* spp., *Cynara* spp., *Daucus carota*, *Desmodium* spp., *Dimocarpus longan*, *Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Elaeis* (e.g. *Elaeis guineensis*, *Elaeis oleifera*), *Eleusine coracana*, *Erianthus* sp., *Eriobotrya japonica*, *Eucalyptus* sp., *Eugenia uniflora*, *Fagopyrum* spp., *Fagus* spp., *Festuca arundinacea*, *Ficus carica*, *Fortunella* spp., *Fragaria* spp., *Ginkgo biloba*, *Glycine* spp. (e.g. *Glycine max*, *Soja hispida* or *Soja max*), *Gossypium hirsutum*, *Helianthus* spp. (e.g. *Helianthus annuus*), *Hemerocallis fulva*, *Hibiscus* spp., *Hordeum* spp. (e.g. *Hordeum vulgare*), *Ipomoea batatas*, *Juglans* spp., *Lactuca sativa*, *Lathyrus* spp., *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* spp., *Luffa acutangula*, *Lupinus* spp., *Luzula sylvatica*, *Lycopersicon* spp. (e.g. *Lycopersicon esculentum*, *Lycopersicon lycopersicum*, *Lycopersicon pyriforme*), *Macrotyloma* spp., *Malus* spp., *Malpighia emarginata*,

Mammea americana, *Mangifera indica*, *Manihot* spp., *Manilkara zapota*, *Medicago sativa*, *Melilotus* spp., *Mentha* spp., *Miscanthus sinensis*, *Momordica* spp., *Morus nigra*, *Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Ornithopus* spp., *Oryza* spp. (e.g. *Oryza sativa*, *Oryza latifolia*), *Panicum miliaceum*, *Panicum virgatum*, *Passiflora edulis*, *Pastinaca sativa*,
5 *Pennisetum* sp., *Persea* spp., *Petroselinum crispum*, *Phalaris arundinacea*, *Phaseolus* spp., *Phleum pratense*, *Phoenix* spp., *Phragmites australis*, *Physalis* spp., *Pinus* spp., *Pistacia vera*, *Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus* spp., *Psidium* spp., *Punica granatum*, *Pyrus communis*, *Quercus* spp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* spp., *Ricinus communis*, *Rubus* spp., *Saccharum* spp., *Salix* sp., *Sambucus* spp., *Secale*
10 *cereale*, *Sesamum* spp., *Sinapis* sp., *Solanum* spp. (e.g. *Solanum tuberosum*, *Solanum integrifolium* or *Solanum lycopersicum*), *Sorghum bicolor*, *Spinacia* spp., *Syzygium* spp., *Tagetes* spp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* spp., *Triticosecale rimpau*, *Triticum* spp. (e.g. *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum* or *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*,
15 *Vaccinium* spp., *Vicia* spp., *Vigna* spp., *Viola odorata*, *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., amongst others.

Detailed description of the invention

I. HARPIN

20 According to a first embodiment, the present invention provides a method for enhancing yield-related traits in plants, comprising modulating expression in a plant of a nucleic acid encoding a Harpin-associated Factor G (hereinafter termed “HpaG”) polypeptide.

A preferred method for modulating (preferably, increasing) expression of a nucleic acid
25 encoding an HpaG polypeptide is by introducing and expressing in a plant a nucleic acid encoding an HpaG polypeptide.

Any reference hereinafter to a “protein useful in the methods of the invention” is taken to mean an HpaG polypeptide as defined herein. Any reference hereinafter to a “nucleic acid useful in
30 the methods of the invention” is taken to mean a nucleic acid capable of encoding such an HpaG polypeptide. The nucleic acid to be introduced into a plant (and therefore useful in performing the methods of the invention) is any nucleic acid encoding the type of protein which will now be described, hereafter also named “*HpaG* nucleic acid” or “*HpaG* gene”.

35 An HpaG polypeptide as defined herein comprises any polypeptide having the following features:

- (i) in increasing order of preference, at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the HpaG polypeptide sequence represented by SEQ ID NO: 2; and
- (ii) an amino acid composition wherein the glycine content ranges from between about 13% and about 25%, the glutamine content ranges from between about 13% and about 20%, the cysteine content ranges from between about 0% and about 1%, the histidine content ranges from between about 0% and about 1%, and wherein tryptophan is absent.
- 10 Preferably, the length of the HpaG polypeptide ranges between about 121 and about 143 amino acids.

Preferably, the HpaG protein also comprises the conserved motif 1 (SEQ ID NO: 3)

G (G/E/D) (N/E) X (Q/R/P) Q (A/S) GX (N/D) G

- 15 wherein X on position 4 may be any amino acid, preferably one of S, N, P, R, or Q, and wherein X on position 9 may be any amino acid, preferably one of Q, E, S, or P; and/or the conserved motif 2 (SEQ ID NO: 4)

(P/A/V) S (P/Q/A) (F/L/Y) TQ (M/A) LM (H/N/Q) IV (G/M) (E/D/Q)

- 20 Optionally, the HpaG protein also has the conserved motif 3:

QGISEKQLDQLL

And/or the conserved motif 4:

ILQAQN

- 25 Furthermore, HpaG polypeptides (at least in their native form) elicit a hypersensitive response in *Arabidopsis thaliana* ecotype Cvi-0 (Kim et al., J. Bacteriol. 185, 3155-3166, 2003).

- Alternatively, the homologue of a HpaG protein has in increasing order of preference at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% overall sequence identity to the amino acid represented by SEQ ID NO: 2, provided that the homologous protein comprises the conserved motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters. Compared to overall

sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered.

The term “domain” and “motif” is as defined in the “definitions” section herein. Specialist
5 databases exist for the identification of domains, for example, SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2002) Nucleic Acids Res 30, 242-244, InterPro (Mulder et al., (2003) Nucl. Acids. Res. 31, 315-318, Prosite (Bucher and Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) ISMB-94; Proceedings 2nd International Conference on
10 Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAIPress, Menlo Park; Hulo et al., Nucl. Acids. Res. 32:D134-D137, (2004), or Pfam (Bateman et al., Nucleic Acids Research 30(1): 276-280 (2002). A set of tools for *in silico* analysis of protein sequences is available on the ExPASy proteomics server (hosted by the Swiss Institute of Bioinformatics (Gasteiger et al., ExPASy: the proteomics
15 server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31:3784-3788(2003)). Domains may also be identified using routine techniques, such as by sequence alignment.

Methods for the alignment of sequences for comparison are well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of
20 Needleman and Wunsch ((1970) J Mol Biol 48: 443-453) to find the global (i.e. spanning the complete sequences) alignment of two sequences that maximizes the number of matches and minimizes the number of gaps. The BLAST algorithm (Altschul et al. (1990) J Mol Biol 215: 403-10) calculates percent sequence identity and performs a statistical analysis of the similarity between the two sequences. The software for performing BLAST analysis is publicly
25 available through the National Centre for Biotechnology Information (NCBI). Homologues may readily be identified using, for example, the ClustalW multiple sequence alignment algorithm (version 1.83), with the default pairwise alignment parameters, and a scoring method in percentage. Global percentages of similarity and identity may also be determined using one of the methods available in the MatGAT software package (Campanella et al., BMC
30 Bioinformatics. 2003 Jul 10;4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.). Minor manual editing may be performed to optimise alignment between conserved motifs, as would be apparent to a person skilled in the art. Furthermore, instead of using full-length sequences for the identification of homologues, specific domains may also be used. The sequence identity values may be determined over
35 the entire nucleic acid or amino acid sequence or over selected domains or conserved motif(s), using the programs mentioned above using the default parameters.

The present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 1, encoding the polypeptide sequence of SEQ ID NO: 2. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any HpaG-encoding nucleic acid or
5 HpaG-like polypeptide as defined herein.

Examples of nucleic acids encoding HpaG polypeptides are given in Table A of Example 1 herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A of Example 1 are example sequences of orthologues and
10 paralogues of the HpaG polypeptide represented by SEQ ID NO: 2, the terms "orthologues" and "paralogues" being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search. Typically, this involves a first BLAST involving BLASTing a query sequence (for example using any of the sequences listed in Table A of Example 1) against any sequence database, such as the publicly available NCBI
15 database. BLASTN or TBLASTX (using standard default values) are generally used when starting from a nucleotide sequence, and BLASTP or TBLASTN (using standard default values) when starting from a protein sequence. The BLAST results may optionally be filtered. The full-length sequences of either the filtered results or non-filtered results are then BLASTed back (second BLAST) against sequences from the organism from which the query sequence is
20 derived (where the query sequence is SEQ ID NO: 1 or SEQ ID NO: 2, the second BLAST would therefore be against *Xanthomonas* sequences). The results of the first and second BLASTs are then compared. A paralogue is identified if a high-ranking hit from the first blast is from the same species as from which the query sequence is derived, a BLAST back then ideally results in the query sequence amongst the highest hits; an orthologue is identified if a
25 high-ranking hit in the first BLAST is not from the same species as from which the query sequence is derived, and preferably results upon BLAST back in the query sequence being among the highest hits.

High-ranking hits are those having a low E-value. The lower the E-value, the more significant
30 the score (or in other words the lower the chance that the hit was found by chance). Computation of the E-value is well known in the art. In addition to E-values, comparisons are also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In the case of large families, ClustalW may be used,
35 followed by a neighbour joining tree, to help visualize clustering of related genes and to identify orthologues and paralogues.

Nucleic acid variants may also be useful in practising the methods of the invention. Examples of such variants include nucleic acids encoding homologues and derivatives of any one of the amino acid sequences given in Table A of Example 1, the terms “homologue” and “derivative” being as defined herein. Also useful in the methods of the invention are nucleic acids encoding homologues and derivatives of orthologues or paralogues of any one of the amino acid sequences given in Table A of Example 1. Homologues and derivatives useful in the methods of the present invention have substantially the same biological and functional activity as the unmodified protein from which they are derived.

Further nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding HpaG polypeptides, nucleic acids hybridising to nucleic acids encoding HpaG polypeptides, and variants of nucleic acids encoding HpaG polypeptides obtained by gene shuffling. The terms hybridising sequence, and gene shuffling are as described herein.

Nucleic acids encoding HpaG polypeptides need not be full-length nucleic acids, since performance of the methods of the invention does not rely on the use of full-length nucleic acid sequences. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a portion of any one of the nucleic acid sequences given in Table A of Example 1, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A of Example 1.

A portion of a nucleic acid may be prepared, for example, by making one or more deletions to the nucleic acid. The portions may be used in isolated form or they may be fused to other coding (or non-coding) sequences in order to, for example, produce a protein that combines several activities. When fused to other coding sequences, the resultant polypeptide produced upon translation may be bigger than that predicted for the protein portion.

Portions useful in the methods of the invention, encode an HpaG polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A of Example 1. Preferably, the portion is a portion of any one of the nucleic acids given in Table A of Example 1, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A of Example 1. Preferably the portion is, in increasing order of preference at least 70, 90, 110, 130 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A of Example 1, or of a nucleic acid encoding an orthologue or

paralogue of any one of the amino acid sequences given in Table A of Example 1. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 1. Preferably, the portion encodes an amino acid sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure. 2, tends to cluster with the group of HpaG polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group.

Another nucleic acid variant useful in the methods of the invention is a nucleic acid capable of hybridising, under reduced stringency conditions, preferably under stringent conditions, with a nucleic acid encoding an HpaG polypeptide as defined herein, or with a portion as defined herein.

According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a nucleic acid capable of hybridizing to any one of the nucleic acids given in Table A of Example 1, or comprising introducing and expressing in a plant a nucleic acid capable of hybridising to a nucleic acid encoding an orthologue, paralogue or homologue of any of the nucleic acid sequences given in Table A of Example 1.

Hybridising sequences useful in the methods of the invention encode an HpaG polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A of Example 1. Preferably, the hybridising sequence is capable of hybridising to any one of the nucleic acids given in Table A of Example 1, or to a portion of any of these sequences, a portion being as defined above, or wherein the hybridising sequence is capable of hybridising to a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A of Example 1. Most preferably, the hybridising sequence is capable of hybridising to a nucleic acid as represented by SEQ ID NO: 1 or to a portion thereof.

Preferably, the hybridising sequence encodes an amino acid sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 2, tends to cluster with the group of HpaG polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group.

Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding HpaG polypeptides as defined above; the term "gene shuffling" being as defined herein.

According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a variant of any one of the nucleic acid sequences given in Table A of Example 1, or comprising introducing and
5 expressing in a plant a variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A of Example 1, which variant nucleic acid is obtained by gene shuffling.

Preferably, the amino acid sequence encoded by the variant nucleic acid obtained by gene
10 shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 2, tends to cluster with the group of HpaG polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group.

Furthermore, nucleic acid variants may also be obtained by site-directed mutagenesis.
15 Several methods are available to achieve site-directed mutagenesis, the most common being PCR based methods (Current Protocols in Molecular Biology. Wiley Eds.).

Nucleic acids encoding HpaG polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic
20 environment through deliberate human manipulation. Preferably the HpaG polypeptide-encoding nucleic acid is of prokaryotic origin, preferably from a Gram-negative bacterium possessing a TTSS, further preferably from a plant pathogenic bacterium possessing a TTSS, more preferably from the family of Pseudomonaceae, furthermore preferably from the genus *Xanthomonas*, most preferably the nucleic acid is from *Xanthomonas axonopodis*.

25 Performance of the methods of the invention gives plants having enhanced yield-related traits. In particular performance of the methods of the invention gives plants having increased yield, especially increased biomass and/or increased seed yield relative to control plants. The terms "yield" and "seed yield" are described in more detail in the "definitions" section herein.

30 Reference herein to enhanced yield-related traits is taken to mean an increase in biomass (weight) of one or more parts of a plant, which may include aboveground (harvestable) parts and/or (harvestable) parts below ground. In particular, such harvestable parts are seeds, and performance of the methods of the invention results in plants having increased seed yield
35 relative to the seed yield of suitable control plants.

Taking corn as an example, a yield increase may be manifested as one or more of the following: increase in the number of plants established per hectare or acre, an increase in the number of ears per plant, an increase in the number of rows, number of kernels per row, kernel weight, thousand kernel weight, ear length/diameter, increase in the seed filling rate (which is the number of filled seeds divided by the total number of seeds and multiplied by 100), among others. Taking rice as an example, a yield increase may manifest itself as an increase in one or more of the following: number of plants per hectare or acre, number of panicles per plant, number of spikelets per panicle, number of flowers (florets) per panicle (which is expressed as a ratio of the number of filled seeds over the number of primary panicles), increase in the seed filling rate (which is the number of filled seeds divided by the total number of seeds and multiplied by 100), increase in thousand kernel weight, among others.

The present invention provides a method for increasing yield, especially biomass and/or seed yield of plants, relative to control plants, which method comprises modulating expression, preferably increasing expression, in a plant of a nucleic acid encoding an HpaG polypeptide as defined herein. It should be noted that the observed yield increase is not the result of increased biotic stress resistance.

Since the transgenic plants according to the present invention have increased yield, it is likely that these plants exhibit an increased growth rate (during at least part of their life cycle), relative to the growth rate of control plants at a corresponding stage in their life cycle. Besides the increased yield capacity, an increased efficiency of nutrient uptake may also contribute to the increase in yield. It is observed that the plants according to the present invention show a higher efficiency in nutrient uptake. Increased efficiency of nutrient uptake allows better growth of the plant.

The increased growth rate may be specific to one or more parts of a plant (including seeds), or may be throughout substantially the whole plant. Plants having an increased growth rate may have a shorter life cycle. The life cycle of a plant may be taken to mean the time needed to grow from a mature seed up to the stage where the plant has produced mature seeds, similar to the starting material. This life cycle may be influenced by factors such as early vigour, growth rate, greenness index, flowering time and speed of seed maturation. The increase in growth rate may take place at one or more stages in the life cycle of a plant or during substantially the whole plant life cycle. Increased growth rate during the early stages in the life cycle of a plant may reflect enhanced vigour. The increase in growth rate may alter the harvest cycle of a plant allowing plants to be sown later and/or harvested sooner than would otherwise be possible (a similar effect may be obtained with earlier flowering time). If the

growth rate is sufficiently increased, it may allow for the further sowing of seeds of the same plant species (for example sowing and harvesting of rice plants followed by sowing and harvesting of further rice plants all within one conventional growing period). Similarly, if the growth rate is sufficiently increased, it may allow for the further sowing of seeds of different plants species (for example the sowing and harvesting of corn plants followed by, for example, the sowing and optional harvesting of soybean, potato or any other suitable plant). Harvesting additional times from the same rootstock in the case of some crop plants may also be possible. Altering the harvest cycle of a plant may lead to an increase in annual biomass production per acre (due to an increase in the number of times (say in a year) that any particular plant may be grown and harvested). An increase in growth rate may also allow for the cultivation of transgenic plants in a wider geographical area than their wild-type counterparts, since the territorial limitations for growing a crop are often determined by adverse environmental conditions either at the time of planting (early season) or at the time of harvesting (late season). Such adverse conditions may be avoided if the harvest cycle is shortened. The growth rate may be determined by deriving various parameters from growth curves, such parameters may be: T-Mid (the time taken for plants to reach 50% of their maximal size) and T-90 (time taken for plants to reach 90% of their maximal size), amongst others.

According to a preferred feature of the present invention, performance of the methods of the invention gives plants having an increased growth rate relative to control plants. Therefore, according to the present invention, there is provided a method for increasing the growth rate of plants, which method comprises modulating expression, preferably increasing expression, in a plant of a nucleic acid encoding an HpaG polypeptide as defined herein. It should be noted that the observed increase in growth rate is not the result of biotic stress resistance.

An increase in yield and/or growth rate occurs whether the plant is under non-stress conditions or whether the plant is exposed to various abiotic stresses compared to control plants. Plants typically respond to exposure to abiotic stress by growing more slowly. In conditions of severe stress, the plant may even stop growing altogether. Mild stress on the other hand is defined herein as being any stress to which a plant is exposed which does not result in the plant ceasing to grow altogether without the capacity to resume growth. Mild stress in the sense of the invention leads to a reduction in the growth of the stressed plants of less than 40%, 35% or 30%, preferably less than 25%, 20% or 15%, more preferably less than 14%, 13%, 12%, 11% or 10% or less in comparison to the control plant under non-stress conditions. Due to advances in agricultural practices (irrigation, fertilization, pesticide treatments) severe stresses are not often encountered in cultivated crop plants. As a consequence, the compromised growth induced by mild stress is often an undesirable feature for agriculture. The term "mild

stresses” are the everyday abiotic (environmental) stresses to which a plant is exposed. Abiotic stresses may be due to drought or excess water, anaerobic stress, salt stress, chemical toxicity, oxidative stress and hot, cold or freezing temperatures. The abiotic stress may be an osmotic stress caused by a water stress (particularly due to drought), salt stress, oxidative stress or an ionic stress.

The term “abiotic stress” as defined herein is taken to mean any one or more of: water stress (due to drought or excess water), anaerobic stress, salt stress, temperature stress (due to hot, cold or freezing temperatures), chemical toxicity stress and oxidative stress. According to one aspect of the invention, the abiotic stress is an osmotic stress, selected from water stress, salt stress, oxidative stress and ionic stress. Preferably, the water stress is drought stress. The term salt stress is not restricted to common salt (NaCl), but may be any one or more of: NaCl, KCl, LiCl, MgCl₂, CaCl₂, amongst others.

Another example of abiotic environmental stress is the reduced availability of one or more nutrients that need to be assimilated by the plants for growth and development. Because of the strong influence of nutrition utilization efficiency on plant yield and product quality, a huge amount of fertilizer is poured onto fields to optimize plant growth and quality. Productivity of plants ordinarily is limited by three primary nutrients, phosphorous, potassium and nitrogen, which is usually the rate-limiting element in plant growth of these three. Therefore the major nutritional element required for plant growth is nitrogen (N). It is a constituent of numerous important compounds found in living cells, including amino acids, proteins (enzymes), nucleic acids, and chlorophyll. 1.5% to 2% of plant dry matter is nitrogen and approximately 16% of total plant protein. Thus, nitrogen availability is a major limiting factor for crop plant growth and production (Frink et al. (1999) Proc Natl Acad Sci USA 96(4): 1175-1180), and has as well a major impact on protein accumulation and amino acid composition. Therefore, of great interest are crop plants with an increased yield when grown under nitrogen-limiting conditions.

Biotic stresses are typically those stresses caused by pathogens, such as bacteria, viruses, fungi, nematodes and insects.

In particular, the methods of the present invention may be performed under non-stress conditions or under conditions of drought to give plants having increased yield relative to control plants. As reported in Wang et al. (Planta (2003) 218: 1-14), abiotic stress leads to a series of morphological, physiological, biochemical and molecular changes that adversely affect plant growth and productivity. Drought, salinity, extreme temperatures and oxidative stress are known to be interconnected and may induce growth and cellular damage through

similar mechanisms. Rabbani et al. (Plant Physiol (2003) 133: 1755-1767) describes a particularly high degree of “cross talk” between drought stress and high-salinity stress. For example, drought and/or salinisation are manifested primarily as osmotic stress, resulting in the disruption of homeostasis and ion distribution in the cell. Oxidative stress, which frequently
5 accompanies high or low temperature, salinity or drought stress, may cause denaturing of functional and structural proteins. As a consequence, these diverse environmental stresses often activate similar cell signalling pathways and cellular responses, such as the production of stress proteins, up-regulation of anti-oxidants, accumulation of compatible solutes and growth arrest.

10

The term “non-stress” conditions as used herein are those environmental conditions that allow optimal growth of plants. Persons skilled in the art are aware of normal soil conditions and climatic conditions for any given location.

15

Performance of the methods of the invention gives plants, grown under non-stress conditions or under drought stress conditions, increased yield relative to suitable control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under non-stress conditions or under drought conditions, which method comprises increasing expression in a plant of a nucleic acid
20 encoding an HpaG polypeptide.

20

Furthermore, performance of the methods of the invention gives plants grown under conditions of nutrient deficiency, particularly under conditions of nitrogen deficiency, increased yield relative to control plants grown under comparable conditions. Therefore, according to the
25 present invention, there is also provided a method for increasing yield in plants grown under conditions of nutrient deficiency, which method comprises increasing expression in a plant of a nucleic acid encoding an HpaG polypeptide.

25

Performance of the methods of the invention also gives plants having increased plant vigour relative to control plants, particularly during the early stages of plant development (typically three, four weeks post germination in the case of rice and maize, but this will vary from species to species) leading to early vigour. Therefore, according to the present invention, there is provided a method for increasing the plant early vigour, which method comprises modulating, preferably increasing, expression in a plant of a nucleic acid encoding a HpaG polypeptide.
30 Preferably the increase in seedling vigour is achieved by expressing the nucleic acid encoding the HpaG polypeptide under the control of a shoot specific promoter. There is also provided a method for producing plants having early vigour relative to control plants, which method
35

30

35

comprises modulating, preferably increasing, expression in a plant of a nucleic acid encoding a HpaG polypeptide.

Early vigour may also result from increased plant fitness due to, for example, the plants being better adapted to their environment (i.e. optimizing the use of energy resources and partitioning between shoot and root). Plants having early vigour also show increase seedling survival and a better establishment of the crop, which often results in highly uniform fields (with the crop growing in uniform manner, i.e. with the majority of plants reaching the various stages of development at substantially the same time), and often better and higher yield. Therefore, early vigour may be determined by measuring various factors, such as thousand kernel weight, percentage germination, percentage emergence, seedling growth, seedling height, root length, root and shoot biomass and many more.

The present invention encompasses plants or parts thereof (including seeds) obtainable by the methods according to the present invention. The plants or parts thereof comprise a nucleic acid transgene encoding an HpaG polypeptide as defined above.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding HpaG polypeptides. The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming into plants and suitable for expression of the gene of interest in the transformed cells. The invention also provides use of a gene construct as defined herein in the methods of the invention.

More specifically, the present invention provides a construct comprising:

- (a) a nucleic acid encoding an HpaG polypeptide as defined above;
- (b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
- (c) a transcription termination sequence.

Preferably, the HpaG encoding nucleic acid is

- (i) a nucleic acid as presented by SEQ ID NO: 1 or the complement thereof,
- (ii) a nucleic acid encoding an HpaG polypeptide as defined above.

The term "control sequence" and "termination sequence" are as defined herein.

Plants are transformed with a vector comprising any of the nucleic acids described above. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells containing the sequence of interest. The sequence of interest is operably linked to one or more control sequences (at least to a promoter).

Advantageously, any type of promoter, whether natural or synthetic, may be used to drive expression of the nucleic acid sequence. A constitutive promoter or a green tissue specific promoter is particularly useful in the methods. See the "Definitions" section herein for definitions of the various promoter types.

Preferably, the *HpaG* nucleic acid or variant thereof is operably linked to a constitutive promoter. A preferred constitutive promoter is one that is also substantially ubiquitously expressed. Further preferably the promoter is derived from a plant, more preferably a monocotyledonous plant. Most preferred is use of a GOS2 promoter (from rice) (SEQ ID NO: 5). It should be clear that the applicability of the present invention is not restricted to the *HpaG* nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to expression of a *HpaG* nucleic acid when driven by a GOS2 promoter. Examples of other constitutive promoters which may also be used to drive expression of an *HpaG* nucleic acid are shown in Table 2a in the Definitions section herein.

Preferably, the consecutive promoter is of medium strength and has weaker activity than the CaMV 35S promoter.

Alternatively, the *HpaG* nucleic acid or variant thereof is operably linked to a green tissue-specific promoter. A green tissue-specific promoter as defined herein is a promoter that is transcriptionally active predominantly in green tissue, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts. The green tissue-specific promoter is preferably a protochlorophyllid reductase promoter, more preferably the protochlorophyllid reductase promoter represented by a nucleic acid sequence substantially similar to SEQ ID NO: 6, most preferably the promoter is as represented by SEQ ID NO: 6. It should be clear that the applicability of the present invention is not restricted to the *HpaG* encoding nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to expression of such a *HpaG* encoding nucleic acid when driven by a protochlorophyllid reductase promoter. Examples of other green tissue-specific promoters which may also be used to perform the methods of the invention are shown in the definitions section herein.

For the identification of functionally equivalent promoters, the promoter strength and/or expression pattern of a candidate promoter may be analysed for example by operably linking the promoter to a reporter gene and assaying the expression level and pattern of the reporter gene in various tissues of the plant. Suitable well-known reporter genes include for example beta-glucuronidase or beta galactosidase. The promoter activity is assayed by measuring the enzymatic activity of the beta-glucuronidase or beta-galactosidase. The promoter strength and/or expression pattern may then be compared to that of a reference promoter (such as the one used in the methods of the present invention). Alternatively, promoter strength may be assayed by quantifying mRNA levels or by comparing mRNA levels of the nucleic acid used in the methods of the present invention, with mRNA levels of housekeeping genes such as 18S rRNA, using methods known in the art, such as Northern blotting with densitometric analysis of autoradiograms, quantitative real-time PCR or RT-PCR (Heid et al., 1996 Genome Methods 6: 986-994). Generally a "weak promoter" refers to a promoter that drives expression of a coding sequence at a low level. By "low level" is intended at levels of about 1/10,000 transcripts to about 1/100,000 transcripts, to about 1/500,000 transcripts per cell. Conversely, a "strong promoter" drives expression of a coding sequence at high level, or at about 1/10 transcripts to about 1/100 transcripts to about 1/1,000 transcripts per cell.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Additional regulatory elements may include transcriptional as well as translational enhancers. Those skilled in the art will be aware of terminator and enhancer sequences that may be suitable for use in performing the invention. Such sequences would be known or may readily be obtained by a person skilled in the art.

An intron sequence may also be added to the 5' untranslated region (UTR) or in the coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, Mol. Cell Biol. 8:4395-4405 (1988); Callis et al., Genes Dev. 1:1183-1200 (1987)). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. For general information, see The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

Other control sequences (besides promoter, enhancer, silencer, intron sequences, 3'UTR and/or 5'UTR regions) may be protein and/or RNA stabilizing elements. Such sequences

would be known or may readily be obtained by a person skilled in the art. Furthermore, the codon usage of the coding sequence to be inserted on the construct may be optimised with reference to the host cell into which the construct will be introduced. While the genetic code is degenerated, organisms tend to use a particular codon for an amino acid more than other
5 codons for that same amino acid. Tables with preferred codon usage for various organisms are known in the art.

The genetic constructs of the invention may further include an origin of replication sequence that is required for maintenance and/or replication in a specific cell type. One example is when
10 a genetic construct is required to be maintained in a bacterial cell as an episomal genetic element (e.g. plasmid or cosmid molecule). Preferred origins of replication include, but are not limited to, the f1-ori and colE1.

For the detection of the successful transfer of the nucleic acid sequences as used in the
15 methods of the invention and/or selection of transgenic plants comprising these nucleic acids, it is advantageous to use marker genes (or reporter genes). Therefore, the genetic construct may optionally comprise a selectable marker gene. Selectable markers are described in more detail in the "definitions" section herein.

20 It is known that upon stable or transient integration of nucleic acids into plant cells, only a minority of the cells takes up the foreign DNA and, if desired, integrates it into its genome, depending on the expression vector used and the transfection technique used. To identify and select these integrants, a gene coding for a selectable marker (such as the ones described
25 above) is usually introduced into the host cells together with the gene of interest. These markers can for example be used in mutants in which these genes are not functional by, for example, deletion by conventional methods. Furthermore, nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector that comprises the sequence encoding the polypeptides of the invention or used in the methods of the invention, or else in a separate vector. Cells which have been stably transfected with the introduced
30 nucleic acid can be identified for example by selection (for example, cells which have integrated the selectable marker survive whereas the other cells die).

Since the marker genes, particularly genes for resistance to antibiotics and herbicides, are no longer required or are undesired in the transgenic host cell once the nucleic acids have been
35 introduced successfully, the process according to the invention for introducing the nucleic acids advantageously employs techniques which enable the removal or excision of these marker genes. One such a method is what is known as co-transformation. The co-

transformation method employs two vectors simultaneously for the transformation, one vector bearing the nucleic acid according to the invention and a second bearing the marker gene(s). A large proportion of transformants receives or, in the case of plants, comprises (up to 40% or more of the transformants), both vectors. In case of transformation with *Agrobacteria*, the transformants usually receive only a part of the vector, i.e. the sequence flanked by the T-DNA, which usually represents the expression cassette. The marker genes can subsequently be removed from the transformed plant by performing crosses. In another method, marker genes integrated into a transposon are used for the transformation together with desired nucleic acid (known as the Ac/Ds technology). The transformants can be crossed with a transposase source or the transformants are transformed with a nucleic acid construct conferring expression of a transposase, transiently or stable. In some cases (approx. 10%), the transposon jumps out of the genome of the host cell once transformation has taken place successfully and is lost. In a further number of cases, the transposon jumps to a different location. In these cases the marker gene must be eliminated by performing crosses. In microbiology, techniques were developed which make possible, or facilitate, the detection of such events. A further advantageous method relies on what is known as recombination systems; whose advantage is that elimination by crossing can be dispensed with. The best-known system of this type is what is known as the Cre/lox system. Cre1 is a recombinase that removes the sequences located between the loxP sequences. If the marker gene is integrated between the loxP sequences, it is removed once transformation has taken place successfully, by expression of the recombinase. Further recombination systems are the HIN/HIX, FLP/FRT and REP/STB system (Tribble et al., J. Biol. Chem., 275, 2000: 22255-22267; Velmurugan et al., J. Cell Biol., 149, 2000: 553-566). A site-specific integration into the plant genome of the nucleic acid sequences according to the invention is possible. Naturally, these methods can also be applied to microorganisms such as yeast, fungi or bacteria.

The invention also provides a method for the production of transgenic plants having enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding an HpaG polypeptide as defined hereinabove.

More specifically, the present invention provides a method for the production of transgenic plants having increased enhanced yield-related traits, particularly increased biomass and/or seed yield, which method comprises:

- (i) introducing and expressing in a plant or plant cell an HpaG polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

The nucleic acid of (i) may be any of the nucleic acids capable of encoding an HpaG polypeptide as defined herein.

The nucleic acid may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of a plant). According to a preferred feature of the present invention, the nucleic acid is preferably introduced into a plant by transformation. The term "transformation" is described in more detail in the "definitions" section herein.

The genetically modified plant cells can be regenerated via all methods with which the skilled worker is familiar. Suitable methods can be found in the abovementioned publications by S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant. To select transformed plants, the plant material obtained in the transformation is, as a rule, subjected to selective conditions so that transformed plants can be distinguished from untransformed plants. For example, the seeds obtained in the above-described manner can be planted and, after an initial growing period, subjected to a suitable selection by spraying. A further possibility consists in growing the seeds, if appropriate after sterilization, on agar plates using a suitable selection agent so that only the transformed seeds can grow into plants. Alternatively, the transformed plants are screened for the presence of a selectable marker such as the ones described above.

Following DNA transfer and regeneration, putatively transformed plants may also be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed and homozygous second-generation (or T2) transformants selected, and the T2 plants may then further be propagated through classical breeding techniques.

The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

5

The present invention clearly extends to any plant cell or plant produced by any of the methods described herein, and to all plant parts and propagules thereof. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced by the parent in the methods according to the invention.

10

15

The invention also includes host cells containing an isolated nucleic acid encoding an HpaG polypeptide as defined hereinabove. Preferred host cells according to the invention are plant cells. Host plants for the nucleic acids or the vector used in the method according to the invention, the expression cassette or construct or vector are, in principle, advantageously all plants, which are capable of synthesizing the polypeptides used in the inventive method.

20

The methods of the invention are advantageously applicable to any plant.

25

Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs. According to a preferred embodiment of the present invention, the plant is a crop plant. Examples of crop plants include soybean, sunflower, canola, alfalfa, rapeseed, cotton, tomato, potato and tobacco. Further preferably, the plant is a monocotyledonous plant. Examples of monocotyledonous plants include sugarcane. More preferably the plant is a cereal. Examples of cereals include rice, maize, wheat, barley, millet, triticale, rye, sorghum and oats.

30

The invention also extends to harvestable parts of a plant such as, but not limited to seeds, leaves, fruits, flowers, stems, rhizomes, tubers and bulbs. The invention furthermore relates to products derived, preferably directly derived, from a harvestable part of such a plant, such as dry pellets or powders, oil, fat and fatty acids, starch or proteins.

35

According to a preferred feature of the invention, the modulated expression is increased expression. Methods for increasing expression of nucleic acids or genes, or gene products, are well documented in the art and include, for example, overexpression driven by appropriate

promoters, the use of transcription enhancers or translation enhancers. Isolated nucleic acids which serve as promoter or enhancer elements may be introduced in an appropriate position (typically upstream) of a non-heterologous form of a polynucleotide so as to upregulate expression. For example, endogenous promoters may be altered in vivo by mutation, deletion, and/or substitution (see, Kmiec, U.S. Pat. No. 5,565,350; Zarling et al., PCT/US93/03868), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

The present invention also encompasses use of nucleic acids encoding HpaG polypeptides as described herein and use of these HpaG polypeptide in enhancing any of the aforementioned yield-related traits in plants.

The methods according to the present invention result in plants having enhanced yield-related traits, as described hereinbefore. These traits may also be combined with other economically advantageous traits, such as further yield-enhancing traits, tolerance to other abiotic and biotic stresses, traits modifying various architectural features and/or biochemical and/or physiological features.

II. SNF2

According to a first embodiment, the present invention provides a method for enhancing yield-related traits in plants relative to control plants, comprising increasing expression in a plant of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide.

A preferred method for increasing expression of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide is by introducing and expressing in a plant a nucleic acid sequence encoding a SWI2/SNF2 polypeptide.

Any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean an SWI2/SNF2 polypeptide as defined herein. Any reference hereinafter to a "nucleic acid sequence useful in the methods of the invention" is taken to mean a nucleic acid sequence

capable of encoding such an SWI2/SNF2 polypeptide. The nucleic acid sequence to be introduced into a plant (and therefore useful in performing the methods of the invention) is any nucleic acid sequence encoding the type of protein, which will now be described, hereafter also named "SWI2/SNF2 nucleic acid sequence" or "SWI2/SNF2 gene".

5

An "SWI2/SNF2 polypeptide" as defined herein refers to any polypeptide which comprises an ATPase domain comprising from N-terminus to C-terminus at least five, preferably six, more preferably seven, most preferably eight of the following motifs:

- 10 (i) Motif I LADDMGLGK(T/S), as represented by SEQ ID N0: 103 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif I;
- (ii) Motif Ia L(L/V/I)(V/I/L)(A/C)P(T/M/V)S(V/I/L)(V/I/L)XNW, as represented by SEQ ID N0: 104 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Ia;
- 15 (iii) Motif II DEAQ(N/A/H)(V/I/L)KN, as represented by SEQ ID N0: 105 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif II;
- 20 (iv) Motif III A(L/M)TGTPXEN, as represented by SEQ ID N0: 106 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif III;
- (v) Motif IV (L/I)XF(T/S)Q(F/Y), as represented by SEQ ID N0: 107 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif IV;
- 25 (vi) Motif V S(L/V)KAGG(V/T/L)G(L/I)(N/T)LTXA(N/S/T)HV, as represented by SEQ ID N0: 108 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif V;
- 30 (vii) Motif Va DRWWNPAVE, as represented by SEQ ID N0: 109 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Va; and
- (viii) Motif VI QA(T/S)DR(A/T/V)(F/Y)R(I/L)GQ, as represented by SEQ ID N0: 110 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif VI,
- 35

where X in Motif Ia, Motif III, Motif IV, and Motif V, is any amino acid.

Alternatively or additionally, an "SWI2/SNF2 polypeptide" as defined herein refers to any polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7 (described in Flaus *et al.* (2006), *supra*), tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30, rather than with any other SWI2/SNF2 clade.

Alternatively or additionally, an "SWI2/SNF2 polypeptide" as defined herein refers to any polypeptide sequence comprising an ATPase domain having in increasing order of preference at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the ATPase domain as represented by SEQ ID NO: 111, comprised in SEQ ID NO: 30.

Alternatively or additionally, an "SWI2/SNF2 polypeptide" as defined herein refers to any polypeptide having in increasing order of preference at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the SWI2/SNF2 polypeptide as represented by SEQ ID NO: 30 or to any of the polypeptide sequences given in Table E herein.

The terms "domain" and "motif" are defined in the "definitions" section herein. Specialist databases exist for the identification of domains, for example, SMART (Schultz *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864; Letunic *et al.* (2002) *Nucleic Acids Res* 30, 242-244), InterPro (Mulder *et al.*, (2003) *Nucl. Acids. Res.* 31, 315-318, Prosite (Bucher and Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) *ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology*. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAI Press, Menlo Park; Hulo *et al.*, (2004) *Nucl. Acids. Res.* 32: D134-D137), or Pfam (Bateman *et al.*, (2002) *Nucleic Acids Research* 30(1): 276-280). A set of tools for in silico analysis of protein sequences is available on the ExPASy proteomics server (hosted by the Swiss Institute of Bioinformatics (Gasteiger *et al.*, (2003) *ExPASy: the proteomics server for in-depth protein knowledge and analysis*, *Nucleic Acids Res* 31: 3784-3788). Domains may also be identified using routine techniques, such as by sequence alignment. Analysis of the polypeptide sequence of SEQ ID NO: 30 is presented below in Examples 9 and 11.

Methods for the alignment of sequences for comparison are well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of Needleman and Wunsch ((1970) J Mol Biol 48: 443-453) to find the global (i.e. spanning the complete sequences) alignment of two sequences that maximizes the number of matches and minimizes the number of gaps. The BLAST algorithm (Altschul et al. (1990) J Mol Biol 215: 403-10) calculates percent sequence identity and performs a statistical analysis of the similarity between the two sequences. The software for performing BLAST analysis is publicly available through the National Centre for Biotechnology Information (NCBI). Homologues may readily be identified using, for example, the ClustalW multiple sequence alignment algorithm (version 1.83), with the default pairwise alignment parameters, and a scoring method in percentage. Global percentages of similarity and identity may also be determined using one of the methods available in the MatGAT software package (Campanella et al., BMC Bioinformatics. 2003 Jul 10;4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.). Minor manual editing may be performed to optimise alignment between conserved motifs, as would be apparent to a person skilled in the art. Furthermore, instead of using full-length sequences for the identification of homologues, specific domains may also be used. The sequence identity values, which are indicated below in Example 3 as a percentage were determined over the entire nucleic acid or polypeptide sequence (Table F herein), and/or over selected domains (such as the ATPase domain as represented by SEQ ID NO: 111, comprised in SEQ ID NO: 30; Table F1 herein) or conserved motif(s), using the programs mentioned above using the default parameters.

The present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 29, encoding the polypeptide sequence of SEQ ID NO: 30. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any SWI2/SNF2-encoding nucleic acid sequence or SWI2/SNF2 polypeptides as defined herein.

Examples of nucleic acid sequences encoding plant SWI2/SNF2 polypeptides are given in Table E of Example 8 herein. Such nucleic acid sequences are useful in performing the methods of the invention. The polypeptide sequences given in Table E of Example 8 are example sequences of orthologues and paralogues of the SWI2/SNF2 polypeptides represented by SEQ ID NO: 30, the terms "orthologues" and "paralogues" being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search. Typically, this involves a first BLAST involving BLASTing a query sequence (for example using any of the sequences listed in Table E of Example 8) against any sequence database, such as the publicly available NCBI database. BLASTN or

TBLASTX (using standard default values) are generally used when starting from a nucleotide sequence, and BLASTP or TBLASTN (using standard default values) when starting from a protein sequence. The BLAST results may optionally be filtered. The full-length sequences of either the filtered results or non-filtered results are then BLASTed back (second BLAST) against sequences from the organism from which the query sequence is derived (where the query sequence is SEQ ID NO: 29 or SEQ ID NO: 30, the second BLAST would therefore be against *Synechocystis* sequences). The results of the first and second BLASTs are then compared. A paralogue is identified if a high-ranking hit from the first blast is from the same species as from which the query sequence is derived, a BLAST back then ideally results in the query sequence amongst the highest hits; an orthologue is identified if a high-ranking hit in the first BLAST is not from the same species as from which the query sequence is derived, and preferably results upon BLAST back in the query sequence being among the highest hits.

High-ranking hits are those having a low E-value. The lower the E-value, the more significant the score (or in other words the lower the chance that the hit was found by chance). Computation of the E-value is well known in the art. In addition to E-values, comparisons are also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In the case of large families, ClustalW may be used, followed by a neighbour joining tree, to help visualize clustering of related genes and to identify orthologues and paralogues (see Figure 7).

Nucleic acid variants may also be useful in practising the methods of the invention. Examples of such variants include nucleic acid sequences encoding homologues and derivatives of any one of the polypeptide sequences given in Table E of Example 8, the terms "homologue" and "derivative" being as defined herein. Also useful in the methods of the invention are nucleic acid sequences encoding homologues and derivatives of orthologues or paralogues of any one of the polypeptide sequences given in Table E of Example 8. Homologues and derivatives useful in the methods of the present invention have substantially the same biological and functional activity as the unmodified protein from which they are derived.

Further nucleic acid variants useful in practising the methods of the invention include portions of nucleic acid sequences encoding SWI2/SNF2 polypeptides, nucleic acid sequences hybridising to nucleic acid sequences encoding SWI2/SNF2 polypeptides, splice variants of nucleic acid sequences encoding SWI2/SNF2 polypeptides, allelic variants of nucleic acid sequences encoding SWI2/SNF2 polypeptides, and variants of nucleic acid sequences

encoding SWI2/SNF2 polypeptides obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

5 Nucleic acid sequences encoding SWI2/SNF2 polypeptides need not be full-length nucleic acid sequences, since performance of the methods of the invention does not rely on the use of full-length nucleic acid sequences. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a portion of any one of the nucleic acid sequences given in Table E of Example 8, or a portion of a nucleic acid sequence encoding an orthologue, paralogue or homologue of any of the
10 polypeptide sequences given in Table E of Example 8.

A portion of a nucleic acid sequence may be prepared, for example, by making one or more deletions to the nucleic acid sequence. The portions may be used in isolated form or they may be fused to other coding (or non-coding) sequences in order to, for example, produce a protein
15 that combines several activities. When fused to other coding sequences, the resultant polypeptide produced upon translation may be bigger than that predicted for the protein portion.

Portions useful in the methods of the invention, encode SWI2/SNF2 polypeptides as defined
20 herein, and have substantially the same biological activity (i.e., enhancing yield-related traits) as the polypeptide sequences given in Table E of Example 8. Preferably, the portion is a portion of any one of the nucleic acid sequences given in Table E of Example 8, or is a portion of a nucleic acid sequence encoding an orthologue or paralogue of any one of the polypeptide sequences given in Table E of Example 8. Preferably the portion is, in increasing order of
25 preference at least 1000, 1100, 1200, 1300 or 1400 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table E of Example 8, or of a nucleic acid sequence encoding an orthologue or paralogue of any one of the polypeptide sequences given in Table E of Example 8. Most preferably the portion is a portion of the nucleic acid sequence of SEQ ID NO: 29. Preferably, the portion encodes a
30 polypeptide sequence comprising any one or more of the domains or motifs defined herein. Preferably, the portion encodes a polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

35 Another nucleic acid variant useful in the methods of the invention is a nucleic acid sequence capable of hybridising, under reduced stringency conditions, preferably under stringent

conditions, with a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined herein, or with a portion as defined herein.

According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a nucleic acid sequence capable of hybridizing to any one of the nucleic acid sequences given in Table E of Example 8, or comprising introducing and expressing in a plant a nucleic acid sequence capable of hybridising to a nucleic acid sequence encoding an orthologue, paralogue or homologue of any of the nucleic acid sequences given in Table E of Example 8.

Hybridising sequences useful in the methods of the invention encode a SWI2/SNF2 polypeptide as defined herein, and have substantially the same biological activity (i.e., enhancing yield-related traits) as the polypeptide sequences given in Table E of Example 8. Preferably, the hybridising sequence is capable of hybridising to any one of the nucleic acid sequences given in Table E of Example 8, or to a portion of any of these sequences, a portion being as defined above, or wherein the hybridising sequence is capable of hybridising to a nucleic acid sequence encoding an orthologue or paralogue of any one of the polypeptide sequences given in Table E of Example 8. Most preferably, the hybridising sequence is capable of hybridising to a nucleic acid sequence as represented by SEQ ID NO: 29 or to a portion thereof. Preferably, the hybridising sequence encodes a polypeptide sequence comprising any one or more of the motifs or domains as defined herein. Preferably, the hybridising sequence encodes a polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding a SWI2/SNF2 polypeptide as defined hereinabove, a splice variant being as defined herein.

According to the present invention, there is provided a method for enhancing yield related traits in plants, comprising introducing and expressing in a plant a splice variant of any one of the nucleic acid sequences given in Table E of Example 8, or a splice variant of a nucleic acid sequence encoding an orthologue, paralogue or homologue of any of the polypeptide sequences given in Table E of Example 8.

The splice variants useful in the methods of the present invention have substantially the same biological activity (i.e., enhancing yield-related traits) as the SWI2/SNF2 polypeptide of SEQ ID

NO: 30 and any of the polypeptide sequences depicted in Table E of Example 8. Preferably, the polypeptide sequence encoded by the splice variant comprises any one or more of the motifs or domains as defined herein. Preferably, the polypeptide sequence encoded by the splice variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined hereinabove, an allelic variant being as defined herein.

According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant an allelic variant of any one of the nucleic acid sequences given in Table E of Example 8, or comprising introducing and expressing in a plant an allelic variant of a nucleic acid sequence encoding an orthologue, paralogue or homologue of any of the polypeptide sequences given in Table E of Example 8.

The allelic variants useful in the methods of the present invention have substantially the same biological activity (i.e., enhancing yield-related traits) as the SWI2/SNF2 polypeptide of SEQ ID NO: 30 and any of the polypeptide sequences depicted in Table E of Example 8. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 29 or an allelic variant of a nucleic acid sequence encoding an orthologue or paralogue of SEQ ID NO: 30. Preferably, the polypeptide sequence encoded by the allelic variant comprises any one or more of the motifs or domains as defined herein. Preferably, the polypeptide sequence encoded by the allelic variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

Gene shuffling or directed evolution may also be used to generate variants of nucleic acid sequences encoding SWI2/SNF2 polypeptides as defined above; the term "gene shuffling" being as defined herein.

According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a variant of any one of the

nucleic acid sequences given in Table E of Example 8, or comprising introducing and expressing in a plant a variant of a nucleic acid sequence encoding an orthologue, paralogue or homologue of any of the polypeptide sequences given in Table E of Example 8, which variant nucleic acid sequence is obtained by gene shuffling.

5

The variant nucleic acid sequences obtained by gene shuffling useful in the methods of the present invention have substantially the same biological activity as the SWI2/SNF2 polypeptide of SEQ ID NO: 30 and any of the polypeptide sequences depicted in Table E of Example 8. Preferably, the variant nucleic acid sequence obtained by gene shuffling encodes a polypeptide sequence comprising any one or more of the motifs or domains as defined herein. Preferably, the polypeptide sequence encoded by the variant nucleic acid sequence obtained by gene shuffling, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

10

15

Furthermore, nucleic acid variants may also be obtained by site-directed mutagenesis. Several methods are available to achieve site-directed mutagenesis, the most common being PCR based methods (Current Protocols in Molecular Biology, Wiley Eds.).

20

Nucleic acid sequences encoding SWI2/SNF2 polypeptides may be derived from any natural or artificial source. The nucleic acid sequence may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the SWI2/SNF2 polypeptide-encoding nucleic acid sequence is from a microbial genome, further preferably from archaea (such from as the following phyla: Crenarchaeota, Euryarchaeota (comprising Halobacteria, Methanobacteria, Methanococci, Methanopyri, Archaeoglobi, Thermoplasmata, and Thermococci classes), Korarchaeota, or Nanoarchaeota) or bacteria (such from as the following phyla: Actinobacteria, Aquificae, Bacteroidetes/Chlorobi, Chlamydiae, Chloroflexi, Chrysiogenetes, Cyanobacteria, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Fibrobacteres/Acidobacteria, Firmicutes, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Nitrospirae, Planctomycetes, Proteobacteria, Spirochaetes, Thermodesulfobacteria, Thermomicrobia, Thermotogae, Verrucomicrobia), more preferably from cyanobacteria, such as *Synechocystis* sp., *Nostoc* sp., *Synechococcus* sp., *Prochlorococcus* sp., *Anaebena* sp., *Gloeobacter* sp., or *Thermosynechococcus* sp., more preferably from *Synechocystis* sp., most preferably from *Synechocystis* sp. PCC6803.

25

30

35

Performance of the methods of the invention gives plants having enhanced yield-related traits relative to control plants.

Reference herein to “enhanced yield-related traits” is taken to mean an increase in biomass (weight) of one or more parts of a plant, which may include aboveground (harvestable) parts and/or (harvestable) parts below ground. In particular, such harvestable parts are seeds, and performance of the methods of the invention results in plants having enhanced seed yield relative to control plants.

Taking corn as an example, a yield increase may be manifested as one or more of the following: increase in the number of plants established per hectare or acre, an increase in the number of ears per plant, an increase in the number of rows, number of kernels per row, kernel weight, thousand kernel weight, ear length/diameter, increase in the seed filling rate (which is the number of filled seeds divided by the total number of seeds and multiplied by 100), among others. Taking rice as an example, a yield increase may manifest itself as an increase in one or more of the following: number of plants per hectare or acre, number of panicles per plant, number of spikelets per panicle, number of flowers (florets) per panicle (which is expressed as a ratio of the number of filled seeds over the number of primary panicles), increase in the seed filling rate (which is the number of filled seeds divided by the total number of seeds and multiplied by 100), increase in thousand kernel weight, among others.

The present invention provides a method for enhancing yield-related traits of plants relative to control plants, which method comprises increasing expression in a plant of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined herein. Preferably, enhanced yield-related traits is one or more of: (i) increased number of flowers per panicle; (ii) increased total seed weight per plant; (iii) increased number of (filled) seeds; or (iv) increased harvest index.

Since the transgenic plants according to the present invention have enhanced yield-related traits, it is likely that these plants exhibit an increased growth rate (during at least part of their life cycle), relative to the growth rate of control plants at a corresponding stage in their life cycle. Besides the increased yield capacity, an increased efficiency of nutrient uptake may also contribute to the increase in yield. It is observed that the plants according to the present invention show a higher efficiency in nutrient uptake. Increased efficiency of nutrient uptake allows better growth of the plant, whether the plant is grown under stress or non-stress conditions.

The increased growth rate may be specific to one or more parts of a plant (including seeds), or may be throughout substantially the whole plant. Plants having an increased growth rate may have a shorter life cycle. The life cycle of a plant may be taken to mean the time needed to grow from a dry mature seed up to the stage where the plant has produced dry mature seeds, similar to the starting material. This life cycle may be influenced by factors such as early vigour, growth rate, greenness index, flowering time and speed of seed maturation. The increase in growth rate may take place at one or more stages in the life cycle of a plant or during substantially the whole plant life cycle. Increased growth rate during the early stages in the life cycle of a plant may reflect enhanced vigour. The increase in growth rate may alter the harvest cycle of a plant allowing plants to be sown later and/or harvested sooner than would otherwise be possible (a similar effect may be obtained with earlier flowering time). If the growth rate is sufficiently increased, it may allow for the further sowing of seeds of the same plant species (for example sowing and harvesting of rice plants followed by sowing and harvesting of further rice plants all within one conventional growing period). Similarly, if the growth rate is sufficiently increased, it may allow for the further sowing of seeds of different plants species (for example the sowing and harvesting of corn plants followed by, for example, the sowing and optional harvesting of soybean, potato or any other suitable plant). Harvesting additional times from the same rootstock in the case of some crop plants may also be possible. Altering the harvest cycle of a plant may lead to an increase in annual biomass production per acre (due to an increase in the number of times (say in a year) that any particular plant may be grown and harvested). An increase in growth rate may also allow for the cultivation of transgenic plants in a wider geographical area than their wild-type counterparts, since the territorial limitations for growing a crop are often determined by adverse environmental conditions either at the time of planting (early season) or at the time of harvesting (late season). Such adverse conditions may be avoided if the harvest cycle is shortened. The growth rate may be determined by deriving various parameters from growth curves, such parameters may be: T-Mid (the time taken for plants to reach 50% of their maximal size) and T-90 (time taken for plants to reach 90% of their maximal size), amongst others.

According to a preferred feature of the present invention, performance of the methods of the invention gives plants having an increased growth rate relative to control plants. Therefore, according to the present invention, there is provided a method for increasing the growth rate of plants, which method comprises increasing expression in a plant of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined herein.

An increase in yield and/or growth occurs whether the plant is grown under non-stress conditions or whether the plant is exposed to various stresses compared to control plants.

Plants typically respond to exposure to stress by growing more slowly. In conditions of severe stress, the plant may even stop growing altogether. Mild stress on the other hand is defined herein as being any stress to which a plant is exposed which does not result in the plant ceasing to grow altogether without the capacity to resume growth. Mild stress in the sense of the invention leads to a reduction in the growth of the stressed plants of less than 40%, 35% or 30%, preferably less than 25%, 20% or 15%, more preferably less than 14%, 13%, 12%, 11% or 10% or less in comparison to the control plant grown under non-stress conditions. Due to advances in agricultural practices (irrigation, fertilization, pesticide treatments) severe stresses are not often encountered in cultivated crop plants. As a consequence, the compromised growth induced by mild stress is often an undesirable feature for agriculture. Mild stresses are the everyday biotic and/or abiotic (environmental) stresses to which a plant is exposed. Abiotic stresses may be due to drought or excess water, anaerobic stress, salt stress, chemical toxicity, oxidative stress and hot, cold or freezing temperatures. The abiotic stress may be an osmotic stress caused by a water stress (particularly due to drought), salt stress, oxidative stress or an ionic stress. Biotic stresses are typically those stresses caused by pathogens, such as bacteria, viruses, fungi, nematodes, and insects. The term "non-stress" conditions as used herein are preferably those environmental conditions that do not significantly go beyond the everyday climatic and other abiotic conditions that plants may encounter most preferably those conditions that allow optimal growth of plants. Persons skilled in the art are aware of normal soil conditions and climatic conditions for a given location.

Performance of the methods of the invention gives plants grown under non-stress conditions or under mild drought conditions having enhanced yield-related traits relative to control plants grown under comparable stress conditions. Therefore, according to the present invention, there is provided a method for enhancing yield-related traits in plants grown under non-stress conditions or under mild drought conditions, which method comprises increasing expression in a plant of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined above.

Performance of the methods according to the present invention results in plants grown under abiotic stress conditions having enhanced yield-related traits relative to control plants grown under comparable stress conditions. As reported in Wang *et al.* (Planta (2003) 218: 1-14), abiotic stress leads to a series of morphological, physiological, biochemical and molecular changes that adversely affect plant growth and productivity. Drought, salinity, extreme temperatures and oxidative stress are known to be interconnected and may induce growth and cellular damage through similar mechanisms. For example, drought and/or salinisation are manifested primarily as osmotic stress, resulting in the disruption of homeostasis and ion distribution in the cell. Oxidative stress, which frequently accompanies high or low

temperature, salinity or drought stress may cause denaturation of functional and structural proteins. As a consequence, these diverse environmental stresses often activate similar cell signaling pathways and cellular responses, such as the production of stress proteins, up-regulation of anti-oxidants, accumulation of compatible solutes and growth arrest. Since
5 diverse environmental stresses activate similar pathways, the exemplification of the present invention with drought stress should not be seen as a limitation to drought stress, but more as a screen to indicate the involvement of SWI2/SNF2 polypeptides as defined above, in enhancing yield-related traits relative to control plants grown in comparable stress conditions, in abiotic stresses in general.

10 A particularly high degree of "cross talk" is reported between drought stress and high-salinity stress (Rabbani *et al.* (2003) Plant Physiol 133: 1755-1767). Therefore, it would be apparent that an SWI2/SNF2 polypeptides would, along with their usefulness in enhancing yield-related traits in plants relative to control plants grown under drought stress conditions, also find use in
15 enhancing yield-related traits in plants, relative to control plants grown under various other abiotic stress conditions.

The term "abiotic stress" as defined herein is taken to mean any one or more of: water stress (due to drought or excess water), anaerobic stress, salt stress, temperature stress (due to hot,
20 cold or freezing temperatures), chemical toxicity stress and oxidative stress. According to one aspect of the invention, the abiotic stress is an osmotic stress, selected from water stress, salt stress, oxidative stress and ionic stress. Preferably, the water stress is drought stress. The term salt stress is not restricted to common salt (NaCl), but may be any one or more of: NaCl, KCl, LiCl, MgCl₂, CaCl₂, amongst others.

25 In particular, the enhanced yield-related traits in plants grown under abiotic stress conditions (preferably under drought stress conditions) relative to control plants grown in comparable stress conditions, may include one or more of the following: (i) increased aboveground area; (ii) increased total root biomass; (iii) increased thick root biomass; (iv) increased thin root
30 biomass; (v) increased number of flowers per panicle; (vi) increased seed fill rate; (vii) increased total seed weight per plant; (viii) increased number of (filled) seeds; or (ix) increased harvest index.

Performance of the methods of the invention gives plants having enhanced yield-related traits
35 under abiotic stress conditions relative to control plants grown in comparable stress conditions. Therefore, according to the present invention, there is provided a method for enhanced yield-related traits in plants grown under abiotic stress conditions, which method comprises

increasing expression in a plant of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide. According to one aspect of the invention, the abiotic stress is an osmotic stress, selected from one or more of the following: water stress, salt stress, oxidative stress and ionic stress. Preferably, the water stress is drought stress.

5

Another example of abiotic environmental stress is the reduced availability of one or more nutrients that need to be assimilated by the plants for growth and development. Because of the strong influence of nutrition utilization efficiency on plant yield and product quality, a huge amount of fertilizer is poured onto fields to optimize plant growth and quality. Productivity of plants ordinarily is limited by three primary nutrients, phosphorous, potassium and nitrogen, which is usually the rate-limiting element in plant growth of these three. Therefore the major nutritional element required for plant growth is nitrogen (N). It is a constituent of numerous important compounds found in living cells, including amino acids, proteins (enzymes), nucleic acids, and chlorophyll. 1.5% to 2% of plant dry matter is nitrogen and approximately 16% of total plant protein. Thus, nitrogen availability is a major limiting factor for crop plant growth and production (Frink et al. (1999) Proc Natl Acad Sci USA 96(4): 1175-1180), and has as well a major impact on protein accumulation and amino acid composition. Therefore, of great interest are crop plants with an increased yield when grown under nitrogen-limiting conditions.

10

15

20

The present invention encompasses plants, parts thereof (including seeds), or plant cells obtainable by the methods according to the present invention. The plants, plant parts or plant cells comprise an isolated nucleic acid transgene encoding an SWI2/SNF2 polypeptide as defined above.

25

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acid sequences encoding SWI2/SNF2 polypeptides. The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming into plants and suitable for expression of the gene of interest in the transformed cells. The invention also provides use of a gene construct as defined herein in the methods of the invention.

30

More specifically, the present invention provides a construct comprising:

(d) a nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined above;

(e) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally

35

(f) a transcription termination sequence.

The term “control sequence” and “termination sequence” are as defined herein.

In one embodiment, one of the control sequences of a construct is a tissue-specific promoter, preferably a promoter for expression in young expanding tissues. An example of a tissue-specific promoter for expression in young expanding tissues is a beta-expansin promoter, for example a rice beta-expansin promoter as represented by SEQ ID NO: 112.

Plants are transformed with a vector comprising any of the nucleic acid sequences described above. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells containing the sequence of interest. The sequence of interest is operably linked to one or more control sequences (at least to a promoter).

Advantageously, any type of promoter may be used to drive expression of the nucleic acid sequence. The promoter may be a constitutive promoter, which refers to a promoter that is transcriptionally active during most, but not necessarily all, phases of its growth and development and under most environmental conditions, in at least one cell, tissue or organ. Alternatively, the promoter may be an inducible promoter, i.e. having induced or increased transcription initiation in response to a chemical (for a review see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89-108), environmental or physical stimulus. Another example of an inducible promoter is a stress-inducible promoter, i.e. a promoter activated when a plant is exposed to various stress conditions, or a pathogen-induced promoter.

Additionally or alternatively, the promoter may be an organ-specific or tissue-specific promoter, i.e. one that is capable of preferentially initiating transcription in certain organs or tissues, such as the leaves, roots, seed tissue etc; or the promoter may be a ubiquitous promoter, which is active in substantially all tissues or cells of an organism, or the promoter may be developmentally regulated, thereby being active during certain developmental stages or in parts of the plant that undergo developmental changes. Promoters able to initiate transcription in certain organs or tissues only are referred to herein as “organ-specific” or “tissue-specific” respectively, similarly, promoters able to initiate transcription in certain cells only are referred to herein as “cell-specific”.

In one embodiment, a nucleic acid sequence encoding SWI2/SNF2 polypeptide as defined above, such as the nucleic acid sequence as represented by SEQ ID NO: 29, is operably linked to a tissue-specific promoter, preferably to a promoter capable of preferentially expressing the nucleic acid sequence in young expanding tissues, or in the apical meristem.

Preferably, the promoter capable of preferentially expressing the nucleic acid sequence in young expanding tissues has a comparable expression profile to a beta-expansin promoter. More specifically, the promoter capable of preferentially expressing the nucleic acid sequence in young expanding tissues is a promoter capable of driving expression in the cell expansion zone of a shoot or root. Most preferably, the promoter capable of preferentially expressing the nucleic acid sequence in young expanding tissues is a beta-expansin promoter, for example a rice beta-expansin promoter as represented by SEQ ID NO: 112.

For the identification of functionally equivalent promoters, the promoter strength and/or expression pattern of a candidate promoter may be analysed for example by operably linking the promoter to a reporter gene and assaying the expression level and pattern of the reporter gene in various tissues of the plant. Suitable well-known reporter genes include for example beta-glucuronidase or beta galactosidase. The promoter activity is assayed by measuring the enzymatic activity of the beta-glucuronidase or beta-galactosidase. The promoter strength and/or expression pattern may then be compared to that of a reference promoter (such as the one used in the methods of the present invention). Alternatively, promoter strength may be assayed by quantifying mRNA levels or by comparing mRNA levels of the nucleic acid sequence used in the methods of the present invention, with mRNA levels of housekeeping genes such as 18S rRNA, using methods known in the art, such as Northern blotting with densitometric analysis of autoradiograms, quantitative real-time PCR or RT-PCR (Heid et al., 1996 Genome Methods 6: 986-994). Generally by "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By "low level" is intended at levels of about 1/10,000 transcripts to about 1/100,000 transcripts, to about 1/500,000 transcripts per cell. Conversely, a "strong promoter" drives expression of a coding sequence at high level, or at about 1/10 transcripts to about 1/100 transcripts to about 1/1,000 transcripts per cell.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Additional regulatory elements may include transcriptional as well as translational enhancers. Those skilled in the art will be aware of terminator and enhancer sequences that may be suitable for use in performing the invention. Such sequences would be known or may readily be obtained by a person skilled in the art.

An intron sequence may also be added to the 5' untranslated region (UTR) or in the coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, Mol. Cell Biol. 8:4395-4405 (1988); Callis et al., Genes

Dev. 1:1183-1200 (1987)). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. For general information, see The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

5

Other control sequences (besides promoter, enhancer, silencer, intron sequences, 3'UTR and/or 5'UTR regions) may be protein and/or RNA stabilizing elements. Such sequences would be known or may readily be obtained by a person skilled in the art.

10 The genetic constructs of the invention may further include an origin of replication sequence that is required for maintenance and/or replication in a specific cell type. One example is when a genetic construct is required to be maintained in a bacterial cell as an episomal genetic element (e.g. plasmid or cosmid molecule). Preferred origins of replication include, but are not limited to, the f1-ori and colE1.

15

For the detection of the successful transfer of the nucleic acid sequences as used in the methods of the invention and/or selection of transgenic plants comprising these nucleic acid sequences, it is advantageous to use marker genes (or reporter genes). Therefore, the genetic construct may optionally comprise a selectable marker gene. Selectable markers are
20 described in more detail in the "definitions" section herein.

It is known that upon stable or transient integration of nucleic acid sequences into plant cells, only a minority of the cells takes up the foreign DNA and, if desired, integrates it into its genome, depending on the expression vector used and the transfection technique used. To
25 identify and select these integrants, a gene coding for a selectable marker (such as the ones described above) is usually introduced into the host cells together with the gene of interest. These markers can for example be used in mutants in which these genes are not functional by, for example, deletion by conventional methods. Furthermore, nucleic acid sequences encoding a selectable marker can be introduced into a host cell on the same vector that comprises the
30 sequence encoding the polypeptides of the invention or used in the methods of the invention, or else in a separate vector. Cells which have been stably transfected with the introduced nucleic acid sequence can be identified for example by selection (for example, cells which have integrated the selectable marker survive whereas the other cells die).

35 Since the marker genes, particularly genes for resistance to antibiotics and herbicides, are no longer required or are undesired in the transgenic host cell once the nucleic acid sequences have been introduced successfully, the process according to the invention for introducing the

nucleic acid sequences advantageously employs techniques, which enable the removal or excision of these marker genes. One such a method is what is known as co-transformation. The co-transformation method employs two vectors simultaneously for the transformation, one vector bearing the nucleic acid sequence according to the invention and a second bearing the marker gene(s). A large proportion of transformants receives or, in the case of plants, comprises (up to 40% or more of the transformants), both vectors. In case of transformation with *Agrobacteria*, the transformants usually receive only a part of the vector, i.e. the sequence flanked by the T-DNA, which usually represents the expression cassette. The marker genes can subsequently be removed from the transformed plant by performing crosses. In another method, marker genes integrated into a transposon are used for the transformation together with desired nucleic acid sequence (known as the Ac/Ds technology). The transformants can be crossed with a transposase source or the transformants are transformed with a nucleic acid construct conferring expression of a transposase, transiently or stable. In some cases (approx. 10%), the transposon jumps out of the genome of the host cell once transformation has taken place successfully and is lost. In a further number of cases, the transposon jumps to a different location. In these cases the marker gene must be eliminated by performing crosses. In microbiology, techniques were developed which make possible, or facilitate, the detection of such events. A further advantageous method relies on what is known as recombination systems; whose advantage is that elimination by crossing can be dispensed with. The best-known system of this type is what is known as the Cre/lox system. Cre1 is a recombinase that removes the sequences located between the loxP sequences. If the marker gene is integrated between the loxP sequences, it is removed once transformation has taken place successfully, by expression of the recombinase. Further recombination systems are the HIN/HIX, FLP/FRT and REP/STB system (Tribble et al., J. Biol. Chem., 275, 2000: 22255-22267; Velmurugan et al., J. Cell Biol., 149, 2000: 553-566). A site-specific integration into the plant genome of the nucleic acid sequences according to the invention is possible. Naturally, these methods can also be applied to microorganisms such as yeast, fungi or bacteria.

The invention also provides a method for the production of transgenic plants having enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined hereinabove.

More specifically, the present invention provides a method for the production of transgenic plants having enhanced yield-related traits relative to control plants, which method comprises:

- (i) introducing and expressing in a plant or plant cell a nucleic acid sequence encoding an SWI2/SNF2 polypeptide; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

The nucleic acid sequence may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of a plant). According to a preferred feature of the present invention, the nucleic acid sequence is preferably introduced
5 into a plant by transformation. The term "transformation" is described in more detail in the "definitions" section herein.

The genetically modified plant cells can be regenerated via all methods with which the skilled worker is familiar. Suitable methods can be found in the abovementioned publications by S.D.
10 Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant.
15 To select transformed plants, the plant material obtained in the transformation is, as a rule, subjected to selective conditions so that transformed plants can be distinguished from untransformed plants. For example, the seeds obtained in the above-described manner can be planted and, after an initial growing period, subjected to a suitable selection by spraying. A further possibility consists in growing the seeds, if appropriate after sterilization, on agar plates
20 using a suitable selection agent so that only the transformed seeds can grow into plants. Alternatively, the transformed plants are screened for the presence of a selectable marker such as the ones described above.

Following DNA transfer and regeneration, putatively transformed plants may also be
25 evaluated, for instance using Southern analysis or quantitative PCR, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed and homozygous second-generation (or T2) transformants selected, and the T2 plants may then further be propagated through classical breeding techniques.
30

The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells

transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

The present invention clearly extends to any plant cell or plant produced by any of the methods described herein, and to all plant parts and propagules thereof. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced by the parent in the methods according to the invention.

The invention also includes host cells containing an isolated nucleic acid sequence encoding an SWI2/SNF2 polypeptide as defined hereinabove. Preferred host cells according to the invention are plant cells. Host plants for the nucleic acid sequences or the vector used in the method according to the invention, the expression cassette or construct or vector are, in principle, advantageously all plants, which are capable of synthesizing the polypeptides used in the inventive method.

The methods of the invention are advantageously applicable to any plant.

Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs. According to a preferred embodiment of the present invention, the plant is a crop plant. Examples of crop plants include soybean, sunflower, canola, alfalfa, rapeseed, cotton, tomato, potato and tobacco. Further preferably, the plant is a monocotyledonous plant. Examples of monocotyledonous plants include sugarcane. More preferably the plant is a cereal. Examples of cereals include rice, maize, wheat, barley, millet, rye, triticale, sorghum and oats.

The invention also extends to harvestable parts of a plant such as, but not limited to seeds, leaves, fruits, flowers, stems, rhizomes, tubers and bulbs. The invention furthermore relates to products derived, preferably directly derived, from a harvestable part of such a plant, such as dry pellets or powders, oil, fat and fatty acids, starch or proteins.

Methods for increasing expression of nucleic acid sequences or genes, or gene products, are well documented in the art and include, for example, overexpression driven by appropriate promoters, the use of transcription enhancers or translation enhancers. Isolated nucleic acid sequences which serve as promoter or enhancer elements may be introduced in an appropriate position (typically upstream) of a non-heterologous form of a polynucleotide so as

to upregulate expression. For example, endogenous promoters may be altered in vivo by mutation, deletion, and/or substitution (see, Kmiec, U.S. Pat. No. 5,565,350; Zarling et al., PCT/US93/03868), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

As mentioned above, a preferred method for increasing expression of a nucleic acid sequence encoding an SWI2/SNF2 polypeptide is by introducing and expressing in a plant a nucleic acid sequence encoding an SWI2/SNF2 polypeptide; however the effects of performing the method, i.e. enhancing yield-related traits, may also be achieved using other well known techniques. A description of some of these techniques will now follow.

One such technique is T-DNA activation tagging (Hayashi et al. Science (1992) 1350-1353), which involves insertion of T-DNA, usually containing a promoter (may also be a translation enhancer or an intron), in the genomic region of the gene of interest or 10 kb up- or downstream of the coding region of a gene in a configuration such that the promoter directs expression of the targeted gene. Typically, regulation of expression of the targeted gene by its natural promoter is disrupted and the gene falls under the control of the newly introduced promoter. The promoter is typically embedded in a T-DNA. This T-DNA is randomly inserted into the plant genome, for example, through Agrobacterium infection and leads to modified expression of genes near the inserted T-DNA. The resulting transgenic plants show dominant phenotypes due to modified expression of genes close to the introduced promoter.

The effects of the invention may also be reproduced using the technique of TILLING (Targeted Induced Local Lesions In Genomes); for a description of the same see the "definitions" section.

The effects of the invention may also be reproduced using homologous recombination; for a description of the same see the "definitions" section.

The present invention also encompasses use of nucleic acid sequences encoding SWI2/SNF2 polypeptides as described herein and use of these SWI2/SNF2 polypeptides in enhancing yield-related traits in plants relative to control plants. Preferably, enhanced yield-related traits is one or more of: (i) increased number of flowers per panicle; (ii) increased total seed weight per plant; (iii) increased number of (filled) seeds; or (iv) increased harvest index.

The present invention further encompasses use of nucleic acid sequences encoding SWI2/SNF2 polypeptides as described herein and use of these SWI2/SNF2 polypeptides in enhancing yield-related traits in plants grown under abiotic stress conditions (preferably under drought stress conditions), relative to control plants grown under comparable stress conditions. Preferably, enhanced yield-related traits are one or more of: (i) increased aboveground area; (ii) increased total root biomass; (iii) increased thick root biomass; (iv) increased thin root biomass; (v) increased number of flowers per panicle; (vi) increased seed fill rate; (vii) increased total seed weight per plant; (viii) increased number of (filled) seeds; or (ix) increased harvest index.

Nucleic acid sequences encoding SWI2/SNF2 polypeptides described herein, or the SWI2/SNF2 polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified, which may be genetically linked to a gene encoding an SWI2/SNF2 polypeptide. The genes/nucleic acid sequences or the SWI2/SNF2 polypeptides themselves may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programmes to select plants having enhanced yield-related traits as defined hereinabove in the methods of the invention.

Allelic variants of a gene/nucleic acid sequence encoding an SWI2/SNF2 polypeptide may also find use in marker-assisted breeding programmes. Such breeding programmes sometimes require introduction of allelic variation by mutagenic treatment of the plants, using for example EMS mutagenesis; alternatively, the programme may start with a collection of allelic variants of so called "natural" origin caused unintentionally. Identification of allelic variants then takes place, for example, by PCR. This is followed by a step for selection of superior allelic variants of the sequence in question and which give enhanced yield-related traits. Selection is typically carried out by monitoring growth performance of plants containing different allelic variants of the sequence in question. Growth performance may be monitored in a greenhouse or in the field. Further optional steps include crossing plants in which the superior allelic variant was identified with another plant. This could be used, for example, to make a combination of interesting phenotypic features.

Nucleic acid sequences encoding SWI2/SNF2 polypeptides may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. Such use of nucleic acid sequences encoding an SWI2/SNF2 polypeptide requires only a nucleic acid sequence of at least 15 nucleotides in length. The nucleic acid sequences encoding an SWI2/SNF2 polypeptide may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Sambrook J, Fritsch EF and Maniatis T (1989) *Molecular Cloning, A Laboratory Manual*) of restriction-digested plant genomic DNA may be probed with nucleic acid sequences encoding the SWI2/SNF2 polypeptide. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1: 174-181) in order to construct a genetic map. In addition, the nucleic acid sequences may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the nucleic acid sequence encoding the SWI2/SNF2 polypeptide in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4: 37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

The nucleic acid probes may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In: *Non-mammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, the nucleic acid probes may be used in direct fluorescence in situ hybridisation (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favour use of large clones (several kb to several hundred kb; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods for genetic and physical mapping may be carried out using the nucleic acid sequences. Examples include allele-specific amplification

(Kazazian (1989) J. Lab. Clin. Med 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) Genomics 16:325-332), allele-specific ligation (Landegren et al. (1988) Science 241:1077-1080), nucleotide extension reactions (Sokolov (1990) Nucleic Acid Res. 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) Nat. Genet. 7:22-28) and Happy Mapping (Dear and Cook (1989) Nucleic Acid Res. 17:6795-6807). For these methods, the sequence of a nucleic acid is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

The methods according to the present invention result in plants having enhanced yield-related traits relative to control plants, as described hereinbefore. This trait may also be combined with other economically advantageous traits, such as further yield-enhancing traits (under normal or stress growth conditions), tolerance to other abiotic and biotic stresses, traits modifying various architectural features and/or biochemical and/or physiological features.

Description of figures

The present invention will now be described with reference to the following figures in which:

Fig. 1 shows an alignment of HpaG polypeptides with motifs 1 and 2 indicated in bold and underlined for SEQ ID NO: 2.

Fig. 2 shows a phylogenetic tree with the group of HpaG polypeptides delineated from other bacterial and from plant proteins (the various sequences are indicated by their GenBank accession numbers and/or gi numbers).

Fig. 3 shows the binary vector for increased expression in *Oryza sativa* of an HpaG protein-encoding nucleic acid from *Xanthomonas* under the control of a rice GOS2 promoter (pGOS2).

Fig. 4 details examples of Harpin sequences useful in performing the methods according to the present invention.

Fig. 5 shows a scheme of the structure of SWI2/SNF2 polypeptides useful in performing the methods of the invention. The SWI2/SNF2 polypeptides useful in performing the methods of the invention comprise an N-terminal domain and an ATPase domain, both marked as an open

box. The typical 8 motifs I, Ia, II, III, IV, V, Va and VI comprised in the ATPase domain of the SWI2/SNF2 polypeptides useful in performing the methods of the invention are marked as black vertical lines.

5 **Fig. 6** shows the sequence logo of the ATPase domain of the 149 SWI2/SNF2 SSO1653 subfamily members as in Flaus *et al.*, (2006). The ATPase domain as represented by SEQ ID NO: 111, and comprised in SEQ ID NO: 30, is in accordance with this sequence logo.

10 **Fig. 7** shows an unrooted radial neighbor-joining tree of SWI2/SNF2 polypeptides from numerous SWI2/SNF2 subfamilies (including the 149 SWI2/SNF2 SSO1653 subfamily members) constructed by Flaus *et al.*, (2006). The polypeptide as represented by SEQ ID NO: 30 is comprised within the SSO1653 cluster (circled in the Figure), together with all the archeal and bacterial (collectively called microbial) SWI2/SNF2 polypeptides.

15 **Fig. 8** shows a CLUSTAL W (1;83) multiple sequence alignment of SWI2/SNF2 polypeptides from various microbes, using default values. SWI2/SNF2 polypeptides share sequence conservation essentially in Motifs I, Ia, II, III, IV, V, Va and VI, comprised in the ATPase domain. These are boxed and identified as such. Another feature that is highlighted is the ATPase domain, for example as represented by SEQ ID NO: 111, comprised in SEQ ID NO:
20 30. The ATPase domain is comprised (from N to C-terminus) between the first amino acid residue of Motif 1 and the last amino acid residue at the C-terminus of the SWI2/SNF2 polypeptide. The beginning and the end of the ATPase domain are marked, and the ATPase domain itself is identified using a black block above the aligned polypeptides.

25 **Fig. 9** shows the binary vector for increased expression in *Oryza sativa* of a *Synechocystis* sp. PCC6803 nucleic acid sequence encoding a SWI2/SNF2 polypeptide under the control of a beta-expansin promoter.

30 **Fig. 10** details examples of SNF2 sequences useful in performing the methods according to the present invention.

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration alone. The following examples are not intended to completely define
35 or otherwise limit the scope of the invention.

Example 1: Identification of HpaG sequences

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 1 and/or protein sequences related to SEQ ID NO: 2 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul *et al.* (1990) J. Mol. Biol. 215:403-410; and Altschul *et al.* (1997) Nucleic Acids Res. 25:3389-3402). The program was used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. The polypeptide encoded by SEQ ID NO: 1 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflects the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search.

Table A provides a list of nucleic acid and protein sequences related to the nucleic acid sequence as represented by SEQ ID NO: 1 and the protein sequence represented by SEQ ID NO: 2.

Table A: HpaG-encoding nucleic acid sequences and HpaG polypeptides useful in the methods of the present invention.

Name	Source organism	Nucleic acid SEQ ID NO:	Polypeptide SEQ ID NO:	Status
HpaG	Xanthomonas axonopodis	1	2	Full length
HpaG_T44C	Synthetic construct	7	8	Full length
HpaG-T	Synthetic construct	9	10	Full length
Hpa1	Xanthomonas axonopodis pv. citri str. 306	11	12	Full length
HpaG-N	Synthetic construct	13	14	Full length
HpaG_G	Xanthomonas axonopodis	15	16	Full length
Hrp	Xanthomonas smithii subsp. smithii	17	18	Full length
hypersensitive response- functioning factor A	Xanthomonas oryzae pv. oryzae strain JXOIII	19	20	Full length
Hpa1	Xanthomonas oryzae pv. oryzae	21	22	Full length
Hpa1	Xanthomonas oryzae pv. oryzae	23	24	Full length

hpaGXooc	Xanthomonas oryzae pv. oryzicola	25	26	Full length
Hpa1	Xanthomonas campestris pv. campestris str. ATCC 33913	27	28	Full length

Example 2: Alignment of HpaG polypeptide sequences

Alignment of polypeptide sequences (Figure 1) was performed using the ClustalW programme which is based on the popular Clustal algorithm of progressive alignment (Thompson *et al.* (1997) Nucleic Acids Res 25:4876-4882; Chenna *et al.* (2003). Nucleic Acids Res 31:3497-3500). Default values are for the gap open penalty of 10, for the gap extension penalty of 0,1 and the selected weight matrix is Blosom 62 (if polypeptides are aligned). Minor manual editing was done to further optimise the alignment.

A phylogenetic tree of HpaG polypeptides (Figure 2) was constructed using a neighbour-joining clustering algorithm as provided in the AlignX programme from the Vector NTI (Invitrogen).

Example 3: Calculation of global percentage identity between polypeptide sequences useful in performing the methods of the invention

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the MatGAT (Matrix Global Alignment Tool) software (Campanella et al., BMC Bioinformatics. 2003 4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences). MatGAT software generates similarity/identity matrices for DNA or protein sequences without needing pre-alignment of the data. The program performs a series of pair-wise alignments using the Myers and Miller global alignment algorithm (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosom 62 (for polypeptides), and then places the results in a distance matrix. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line.

Parameters used in the comparison were:

Scoring matrix: Blosom62

First Gap: 12

Extending gap: 2

Results of the software analysis are shown in Table B for the global similarity and identity over the full length of the polypeptide sequences (excluding the partial polypeptide sequences).

Percentage identity is given above the diagonal in bold and percentage similarity is given below the diagonal (normal face).

The percentage identity between the HpaG polypeptide sequences useful in performing the methods of the invention can be as low as 37 % amino acid identity compared to SEQ ID NO: 9.

Table B: MatGAT results for global similarity and identity over the full length of the polypeptide sequences.

	1	2	3	4	5	6	7	8	9	10	11	12
1. SEQ ID NO: 2		99.2	94.0	91.2	91.0	90.2	85.4	66.7	66.7	66.7	59.6	37.7
2. ABK51589	99.2		93.2	90.5	90.2	89.5	84.7	67.4	67.4	67.4	60.3	37.7
3. ABK51587	94.0	93.2		85.4	85.0	92.0	79.6	60.3	60.3	60.3	56.4	33.3
4. AAM35307	92.0	91.2	86.1		82.5	81.8	89.8	70.9	70.9	70.9	61.4	36.6
5. ABK51590	91.0	90.2	90.4	83.2		81.2	76.6	57.4	57.4	57.4	50.7	32.8
6. ABK51588	90.2	89.5	92.0	82.5	89.3		75.2	58.2	58.2	58.2	56.4	33.8
7. ABG36696	89.5	88.7	83.5	92.7	80.5	79.7		70.7	70.7	70.7	58.8	37.0
8. ABJ97680	77.0	77.7	70.5	80.6	67.6	68.3	81.3		100.0	100.0	64.5	35.0
9. AAC95121	77.0	77.7	70.5	80.6	67.6	68.3	81.3	100.0		100.0	64.5	35.0
10. BAD29979	77.0	77.7	70.5	80.6	67.6	68.3	81.3	100.0	100.0		64.5	35.0
11. ABB72197	72.9	73.7	72.8	73.7	68.0	72.8	72.9	72.7	72.7	72.7		34.6
12. AAM40538	51.9	51.9	48.0	49.6	46.3	50.4	50.4	45.3	45.3	45.3	53.6	

Example 4: Cloning and vector construction

Unless otherwise stated, recombinant DNA techniques are performed according to standard protocols described in (Sambrook (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York) or in Volumes 1 and 2 of Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfax (1993) by R.D.D. Croy, published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

The *Xanthomonas* HpaG coding sequence was amplified by PCR from a *Xanthomonas axonopodis* DNA library. The PCR fragment of the expected length was purified and subsequently cloned in a Gateway[®] vector using standard technology. The entry clone comprising SEQ ID NO: 1 was then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA

borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 5) for constitutive expression was located upstream of this Gateway cassette. Alternatively, a green tissue specific promoter, such as the protochlorophyllide reductase promoter (SEQ ID NO: 6), was shown to be equally useful.

After the LR recombination step, the resulting expression vector pGOS2::HpaG was transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

Example 5: Plant transformation

Rice transformation

The *Agrobacterium* containing the expression vector was used to transform *Oryza sativa* plants. Mature dry seeds of the rice japonica cultivar Nipponbare were dehusked. Sterilization was carried out by incubating for one minute in 70% ethanol, followed by 30 minutes in 0.2% HgCl₂, followed by a 6 times 15 minutes wash with sterile distilled water. The sterile seeds were then germinated on a medium containing 2,4-D (callus induction medium). After incubation in the dark for four weeks, embryogenic, scutellum-derived calli were excised and propagated on the same medium. After two weeks, the calli were multiplied or propagated by subculture on the same medium for another 2 weeks. Embryogenic callus pieces were subcultured on fresh medium 3 days before co-cultivation (to boost cell division activity).

Agrobacterium strain LBA4404 containing the expression vector was used for co-cultivation. *Agrobacterium* was inoculated on AB medium with the appropriate antibiotics and cultured for 3 days at 28°C. The bacteria were then collected and suspended in liquid co-cultivation medium to a density (OD₆₀₀) of about 1. The suspension was then transferred to a Petri dish and the calli immersed in the suspension for 15 minutes. The callus tissues were then blotted dry on a filter paper and transferred to solidified, co-cultivation medium and incubated for 3 days in the dark at 25°C. Co-cultivated calli were grown on 2,4-D-containing medium for 4 weeks in the dark at 28°C in the presence of a selection agent. During this period, rapidly growing resistant callus islands developed. After transfer of this material to a regeneration medium and incubation in the light, the embryogenic potential was released and shoots developed in the next four to five weeks. Shoots were excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium from which they were transferred to soil. Hardened shoots were grown under high humidity and short days in a greenhouse.

Approximately 35 independent T0 rice transformants were generated for one construct. The primary transformants were transferred from a tissue culture chamber to a greenhouse. After a quantitative PCR analysis to verify copy number of the T-DNA insert, only single copy transgenic plants that exhibit tolerance to the selection agent were kept for harvest of T1 seed.

5 Seeds were then harvested three to five months after transplanting. The method yielded single locus transformants at a rate of over 50 % (Aldemita and Hodges1996, Chan *et al.* 1993, Hiei *et al.* 1994).

Corn transformation

10 Transformation of maize (*Zea mays*) is performed with a modification of the method described by Ishida et al. (1996) Nature Biotech 14(6): 745-50. Transformation is genotype-dependent in corn and only specific genotypes are amenable to transformation and regeneration. The inbred line A188 (University of Minnesota) or hybrids with A188 as a parent are good sources of donor material for transformation, but other genotypes can be used successfully as well. Ears
15 are harvested from corn plant approximately 11 days after pollination (DAP) when the length of the immature embryo is about 1 to 1.2 mm. Immature embryos are cocultivated with *Agrobacterium tumefaciens* containing the expression vector, and transgenic plants are recovered through organogenesis. Excised embryos are grown on callus induction medium, then maize regeneration medium, containing the selection agent (for example imidazolinone
20 but various selection markers can be used). The Petri plates are incubated in the light at 25 °C for 2-3 weeks, or until shoots develop. The green shoots are transferred from each embryo to maize rooting medium and incubated at 25 °C for 2-3 weeks, until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Wheat transformation

Transformation of wheat is performed with the method described by Ishida et al. (1996) Nature Biotech 14(6): 745-50. The cultivar Bobwhite (available from CIMMYT, Mexico) is commonly used in transformation. Immature embryos are co-cultivated with *Agrobacterium tumefaciens*
30 containing the expression vector, and transgenic plants are recovered through organogenesis. After incubation with *Agrobacterium*, the embryos are grown *in vitro* on callus induction medium, then regeneration medium, containing the selection agent (for example imidazolinone but various selection markers can be used). The Petri plates are incubated in the light at 25 °C for 2-3 weeks, or until shoots develop. The green shoots are transferred from each embryo to
35 rooting medium and incubated at 25 °C for 2-3 weeks, until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Soybean transformation

Soybean is transformed according to a modification of the method described in the Texas A&M patent US 5,164,310. Several commercial soybean varieties are amenable to transformation by this method. The cultivar Jack (available from the Illinois Seed foundation) is commonly used for transformation. Soybean seeds are sterilised for *in vitro* sowing. The hypocotyl, the radicle and one cotyledon are excised from seven-day old young seedlings. The epicotyl and the remaining cotyledon are further grown to develop axillary nodes. These axillary nodes are excised and incubated with *Agrobacterium tumefaciens* containing the expression vector. After the cocultivation treatment, the explants are washed and transferred to selection media. Regenerated shoots are excised and placed on a shoot elongation medium. Shoots no longer than 1 cm are placed on rooting medium until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Rapeseed/canola transformation

Cotyledonary petioles and hypocotyls of 5-6 day old young seedling are used as explants for tissue culture and transformed according to Babic et al. (1998, Plant Cell Rep 17: 183-188). The commercial cultivar Westar (Agriculture Canada) is the standard variety used for transformation, but other varieties can also be used. Canola seeds are surface-sterilized for *in vitro* sowing. The cotyledon petiole explants with the cotyledon attached are excised from the *in vitro* seedlings, and inoculated with *Agrobacterium* (containing the expression vector) by dipping the cut end of the petiole explant into the bacterial suspension. The explants are then cultured for 2 days on MSBAP-3 medium containing 3 mg/l BAP, 3 % sucrose, 0.7 % Phytagar at 23 °C, 16 hr light. After two days of co-cultivation with *Agrobacterium*, the petiole explants are transferred to MSBAP-3 medium containing 3 mg/l BAP, cefotaxime, carbenicillin, or timentin (300 mg/l) for 7 days, and then cultured on MSBAP-3 medium with cefotaxime, carbenicillin, or timentin and selection agent until shoot regeneration. When the shoots are 5 – 10 mm in length, they are cut and transferred to shoot elongation medium (MSBAP-0.5, containing 0.5 mg/l BAP). Shoots of about 2 cm in length are transferred to the rooting medium (MS0) for root induction. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Alfalfa transformation

A regenerating clone of alfalfa (*Medicago sativa*) is transformed using the method of (McKersie et al., 1999 Plant Physiol 119: 839–847). Regeneration and transformation of alfalfa is

genotype dependent and therefore a regenerating plant is required. Methods to obtain regenerating plants have been described. For example, these can be selected from the cultivar Ranglander (Agriculture Canada) or any other commercial alfalfa variety as described by Brown DCW and A Atanassov (1985. Plant Cell Tissue Organ Culture 4: 111-112).

5 Alternatively, the RA3 variety (University of Wisconsin) has been selected for use in tissue culture (Walker et al., 1978 Am J Bot 65:654-659). Petiole explants are cocultivated with an overnight culture of *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie et al., 1999 Plant Physiol 119: 839-847) or LBA4404 containing the expression vector. The explants are cocultivated for 3 d in the dark on SH induction medium containing 288 mg/ L Pro, 53 mg/ L thioproline, 4.35 g/ L K₂SO₄, and 100 µm acetosyringinone. The explants are washed in half-
10 strength Murashige-Skoog medium (Murashige and Skoog, 1962) and plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit *Agrobacterium* growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and
15 50 g/ L sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were transplanted into pots and grown in a greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

20 *Cotton transformation*

Cotton is transformed using *Agrobacterium tumefaciens* according to the method described in US 5,159,135. Cotton seeds are surface sterilised in 3% sodium hypochlorite solution during 20 minutes and washed in distilled water with 500 µg/ml cefotaxime. The seeds are then transferred to SH-medium with 50µg/ml benomyl for germination. Hypocotyls of 4 to 6 days
25 old seedlings are removed, cut into 0.5 cm pieces and are placed on 0.8% agar. An *Agrobacterium* suspension (approx. 10⁸ cells per ml, diluted from an overnight culture transformed with the gene of interest and suitable selection markers) is used for inoculation of the hypocotyl explants. After 3 days at room temperature and lighting, the tissues are transferred to a solid medium (1.6 g/l Gelrite) with Murashige and Skoog salts with B5 vitamins
30 (Gamborg et al., Exp. Cell Res. 50:151-158 (1968)), 0.1 mg/l 2,4-D, 0.1 mg/l 6-furfurylaminopurine and 750 µg/ml MgCL₂, and with 50 to 100 µg/ml cefotaxime and 400-500 µg/ml carbenicillin to kill residual bacteria. Individual cell lines are isolated after two to three months (with subcultures every four to six weeks) and are further cultivated on selective medium for tissue amplification (30°C, 16 hr photoperiod). Transformed tissues are
35 subsequently further cultivated on non-selective medium during 2 to 3 months to give rise to somatic embryos. Healthy looking embryos of at least 4 mm length are transferred to tubes with SH medium in fine vermiculite, supplemented with 0.1 mg/l indole acetic acid, 6

furfurylaminopurine and gibberellic acid. The embryos are cultivated at 30°C with a photoperiod of 16 hrs, and plantlets at the 2 to 3 leaf stage are transferred to pots with vermiculite and nutrients. The plants are hardened and subsequently moved to the greenhouse for further cultivation.

5

Example 6: Phenotypic evaluation procedure

6.1 Evaluation setup

Approximately 35 independent T0 rice transformants were generated. The primary transformants were transferred from a tissue culture chamber to a greenhouse for growing and harvest of T1 seed. Six events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes) and approximately 10 T1 seedlings lacking the transgene (nullizygotes) were selected by monitoring visual marker expression. The transgenic plants and the corresponding nullizygotes were grown side-by-side at random positions. Greenhouse conditions were of shorts days (12 hours light), 28°C in the light and 22°C in the dark, and a relative humidity of 70%.

Four T1 events were further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation but with more individuals per event. From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles.

Drought screen

Plants from six events (T2 seeds) were grown in potting soil under normal conditions until they approached the heading stage. They were then transferred to a "dry" section where irrigation was withheld. Humidity probes were inserted in randomly chosen pots to monitor the soil water content (SWC). When SWC went below certain thresholds, the plants were automatically re-watered continuously until a normal level was reached again. The plants were then re-transferred again to normal conditions. The rest of the cultivation (plant maturation, seed harvest) was the same as for plants not grown under abiotic stress conditions. Growth and yield parameters are recorded as detailed for growth under normal conditions.

Nitrogen use efficiency screen

Rice plants from T2 seeds are grown in potting soil under normal conditions except for the nutrient solution. The pots are watered from transplantation to maturation with a specific nutrient solution containing reduced N nitrogen (N) content, usually between 7 to 8 times less.

The rest of the cultivation (plant maturation, seed harvest) is the same as for plants not grown under abiotic stress. Growth and yield parameters are recorded as detailed for growth under normal conditions.

5 *Salt stress screen*

Plants are grown on a substrate made of coco fibers and argex (3 to 1 ratio). A normal nutrient solution is used during the first two weeks after transplanting the plantlets in the greenhouse. After the first two weeks, 25 mM of salt (NaCl) is added to the nutrient solution, until the plants are harvested. Seed-related parameters were then measured.

10

6.2 Statistical analysis: F-test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F-test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention.

15

The F-test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F-test. A significant F-test value points to a gene effect, meaning that it is not only the mere presence or position of the gene that is causing the differences in phenotype.

20

Because two experiments with overlapping events were carried out, a combined analysis was performed. This is useful to check consistency of the effects over the two experiments, and if this is the case, to accumulate evidence from both experiments in order to increase confidence in the conclusion. The method used was a mixed-model approach that takes into account the

25 multilevel structure of the data (i.e. experiment - event - segregants). P-values were obtained by comparing likelihood ratio test to chi square distributions.

6.3 Parameters measured

Biomass-related parameter measurement

30

From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles.

35

The plant aboveground area (or leafy biomass) was determined by counting the total number of pixels on the digital images from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from the different angles and was converted to a physical surface value expressed in square mm by calibration. Experiments show that the aboveground plant area measured this way correlates

with the biomass of plant parts above ground. The above ground area is the area measured at the time point at which the plant had reached its maximal leafy biomass. The early vigour is the plant (seedling) aboveground area three weeks post-germination. Increase in root biomass is expressed as an increase in total root biomass (measured as maximum biomass of roots observed during the lifespan of a plant); or as an increase in the root/shoot index (measured as the ratio between root mass and shoot mass in the period of active growth of root and shoot).

Early vigour was determined by counting the total number of pixels from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from different angles and was converted to a physical surface value expressed in square mm by calibration. The results described below are for plants three weeks post-germination.

Seed-related parameter measurements

The mature primary panicles were harvested, counted, bagged, barcode-labelled and then dried for three days in an oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. The number of filled seeds was determined by counting the number of filled husks that remained after the separation step. The total seed yield was measured by weighing all filled husks harvested from a plant. Total seed number per plant was measured by counting the number of husks harvested from a plant. Thousand Kernel Weight (TKW) is extrapolated from the number of filled seeds counted and their total weight. The Harvest Index (HI) in the present invention is defined as the ratio between the total seed yield and the above ground area (mm²), multiplied by a factor 10⁶. The total number of flowers per panicle as defined in the present invention is the ratio between the total number of seeds and the number of mature primary panicles. The seed fill rate as defined in the present invention is the proportion (expressed as a %) of the number of filled seeds over the total number of seeds (or florets).

Example 7: Results of the phenotypic evaluation of the transgenic plants

The results of the evaluation of transgenic rice plants expressing an HpaG nucleic acid under non-stress conditions are presented below. An increase was observed for aboveground biomass (AreaMax), emergence vigour (early vigour), total seed yield, number of filled seeds, fill rate, number of flowers per panicle, harvest index, and thousand kernel weight (see table C)

Table C: Results of the measurements for yield increase under non-stress conditions

Parameter	Overall increase (in %)	p-value of F-test
AreaMax	13	0.0000
Early vigour	25	0.0041
Total weight of seeds	30	0.0000
Nr of filled seeds	26	0.0000
Fill rate	9	0.0000
Flowers per panicle	12	0.0371
Harvest Index	18	0.0000
Thousand Kernel Weight	4	0.0000

The results of the evaluation of transgenic rice plants expressing an HpaG nucleic acid under drought-stress conditions are presented hereunder. An increase was observed for total seed weight, number of filled seeds, fill rate, harvest index and thousand-kernel weight (Table D).

5

Table D: Results of the measurements for yield increase under drought stress conditions

Parameter	Overall increase (in %)	p-value of F-test
Total weight of seeds	40	0.0000
Nr of filled seeds	37	0.0000
Fill rate	30	0.0000
Harvest Index	37	0.0000
Thousand Kernel Weight	3	0.0001

Example 8: Identification of sequences related to SEQ ID NO: 29 and SEQ ID NO: 30

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 29 and/or protein sequences related to SEQ ID NO: 30 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul *et al.* (1990) J. Mol. Biol. 215:403-410; and Altschul *et al.* (1997) Nucleic Acids Res. 25:3389-3402). The program was used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. The polypeptide encoded by SEQ ID NO: 29 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflects the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage

identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search.

- 5 **Table E** provides a list of nucleic acid and polypeptide sequences related to the nucleic acid sequence as represented by SEQ ID NO: 29 and the polypeptide sequence represented by SEQ ID NO: 30.

Name	Source organism	NCBI polypeptide accession number	NA SEQ ID NO	AA SEQ ID NO
Synecho_PCC6803_SNF2	Synechocystis sp. PCC 6803 BA000022	NP_442847.1	29	30
Anava_SNF2	Anaebena variabilis ATCC 29413	YP_323780.1	31	32
Archaeon RC-I_SNF2	Uncultured methanogenic archaeon RC-I_SNF2	CAJ35100.1	33	34
Bacce_ATCC10987_SNF2	Bacillus cereus ATCC 10987	AAS44264.1	35	36
Crowa_SNF2	Crocospaera watsonii WH 8501 ctg336	ZP_00516613.1	37	38
Glovi_SNF2	Gloeobacter violaceus PCC 7421	NP_925212	39	40
Lyn_sp_SNF2	Lyngbya sp. PCC 8106	ZP_01622333.1	41	42
Metac_C2A_SNF2	Methanosarcina acetivorans C2A	NP_615162.1	43	44
Methu_JF-1_SNF2	Methanospirillum hungatei JF-1	ABD41401.1	45	46
Metma_Go1_SNF2	Methanosarcina mazei Goe1	NP_633503.1	47	48
Mycbo_SNF2	Mycobacterium bovis BCG Pasteur 1173P2	CAL72108.1	49	50
Myctu_SNF2	Mycobacterium tuberculosis H37Rv	BX842578.1	51	52
Myxxa_DK_SNF2	Myxococcus xanthus DK 1622	YP_635387.1	53	54
Nocfa_IFM 10152_SNF2	Nocardia farcinica IFM 10152	BAD55876.1	55	56
Nodsp_SNF2	Nodularia spumigena	ZP_01629192.1	57	58
Nos_sp_PCC7120_SNF2	Nostoc sp. PCC7120	BAB78256.1	59	60
Nos_sp_PCC7120_SNF2 II	Nostoc sp. PCC 7120	ZP_00106150.1	61	62
Nospu_PCC 73102_SNF2	Nostoc punctiforme PCC 73102	NP_488438	63	64
Pelph_BU-1_SNF2	Pelodictyon phaeoclathratiforme BU-1	ZP_00589405.1	65	66
Proma_CCMP1375_SNF2	Prochlorococcus marinus str. CCMP1375	NP_874441.1	67	68
Proma_MIT 9211_SNF2	Prochlorococcus marinus str. MIT 9211	ZP_01006255.1	69	70
Proma_MIT 9303_SNF2	Prochlorococcus marinus str. MIT 9303	YP_001018833.1	71	72
Proma_MIT9313_SNF2	Prochlorococcus marinus str. MIT 9313	NP_895982.1	73	74
Rho_sp_RHA1_SNF2	Rhodococcus sp. RHA1	ABG93371.1	75	76
Saltr_CNB-440_SNF2	Salinispora tropica CNB-440	ZP_01431310	77	78

Symth_IAM14863_SNF2	Symbiobacterium thermophilum IAM 14863	BAD39642	79	80
Syn_sp_WH 5701_SNF2	Synechococcus sp. WH 5701	ZP_01083591.1	81	82
Syn_sp_BL107_SNF2	Synechococcus sp. BL107	ZP_01469219.1	83	84
Syn_sp_CC9311_SNF2	Synechococcus sp. CC9311	YP_731958.1	85	86
Syn_sp_CC9605_SNF2	Synechococcus sp. CC9605	YP_382805.1	87	88
Syn_sp_CC9902_SNF2	Synechococcus sp. CC9902	YP_378176.1	89	90
Syn_sp_RS9916_SNF2	Synechococcus sp. RS9916	ZP_01471362	91	92
Syn_sp_WH 7805_SNF2	Synechococcus sp. WH 7805	ZP_01125039.1	93	94
Syn_sp_WH 8102_SNF2	Synechococcus sp. WH 8102	NP_898451.1	95	96
Synel_PCC6301_SNF2	Synechococcus elongatus PCC 6301	YP_171376	97	98
Synel_PCC7942_SNF2	Synechococcus elongatus PCC 7942	YP_399891.1	99	100
Theel_BP-1_SNF2	Thermosynechococcus elongatus BP-1	NP_682403.1	101	102

Additional sources of SWI2/SNF2 polypeptides useful in performing the methods of the invention can be found in the supplementary table S1C provided by Flaus *et al.* (2006). The authors scanned 24 complete archeal and 269 bacterial genomes, and identified 149 SWI2/SNF2 of the SSO1653 subfamily type.

5

Example 9: Alignment of SWI2/SNF2 polypeptide sequences

Alignment of polypeptide sequences was performed the Clustal algorithm (1.83) of progressive alignment, using default values (Thompson *et al.* (1997) Nucleic Acids Res 25:4876-4882; Chenna *et al.* (2003). Nucleic Acids Res 31:3497-3500). Results in Figure 8 show that SWI2/SNF2 polypeptides share sequence conservation essentially in Motifs I, Ia, II, III, IV, V, Va and VI (which are boxed), represented as follows:

10

15

20

- (i) Motif I LADDMGLGK(T/S), as represented by SEQ ID N0: 103 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif I;
- (ii) Motif Ia L(L/V/I)(V/I/L)(A/C)P(T/M/V)S(V/I/L)(V/I/L)XNW, as represented by SEQ ID N0: 104 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Ia;
- (iii) Motif II DEAQ(N/A/H)(V/I/L)KN, as represented by SEQ ID N0: 105 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif II;

- (iv) Motif III A(L/M)TGTPXEN, as represented by SEQ ID NO: 106 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif III;
- (v) Motif IV (L/I)XF(T/S)Q(F/Y), as represented by SEQ ID NO: 107 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif IV;
- (vi) Motif V S(L/V)KAGG(V/T/L)G(L/I)(N/T)LTXA(N/S/T)HV, as represented by SEQ ID NO: 108 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif V;
- (vii) Motif Va DRWWNPAVE, as represented by SEQ ID NO: 109 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Va; and
- (viii) Motif VI QA(T/S)DR(A/T/V)(F/Y)R(I/L)GQ, as represented by SEQ ID NO: 110 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif VI,

where X in Motif Ia, Motif III, Motif IV, and Motif V, is any amino acid.

These eight motifs are comprised within the ATPase domain. The ATPase domain is comprised (from N-terminus to C-terminus) between the first amino acid residue of Motif 1 and the last amino acid residue at the C-terminus of the SWI2/SNF2 polypeptide. The beginning and the end of the ATPase domain are marked in Figure 8, and the ATPase domain itself is identified using a black block above the aligned polypeptides. An example of an ATPase domain is the ATPase domain of SEQ ID NO: 30, represented by SEQ ID NO: 111.

The sequence logo of the ATPase domain of the 149 SWI2/SNF2 SSO1653 subfamily members is presented in Flaus *et al.*, (2006), and shown in Figure 6. Sequence logos are a graphical representation of an amino acid or nucleic acid multiple sequence alignment. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence. The algorithm (WebLogo) to produce such logos is available at the server of the University of California, Berkeley. The

ATPase domain as represented by SEQ ID NO: 111, and comprised in SEQ ID NO: 30, is in accordance with the sequence logo as represented in Figure 6.

An unrooted radial neighbor-joining tree of SWI2/SNF2 polypeptides from numerous SWI2/SNF2 subfamilies (including SSO1653) was constructed by Flaus *et al.*, (2006), as shown in Figure 7. The polypeptide as represented by SEQ ID NO: 30 is comprised within the SSO1653 cluster (circled in the Figure), together with all the archeal and bacterial (collectively called microbial) SWI2/SNF2 polypeptides.

Example 10: Calculation of global percentage identity between polypeptide sequences useful in performing the methods of the invention

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the MatGAT (Matrix Global Alignment Tool) software (BMC Bioinformatics. 2003 4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. Campanella JJ, Bitincka L, Smalley J; software hosted by Ledion Bitincka). MatGAT software generates similarity/identity matrices for DNA or protein sequences without needing pre-alignment of the data. The program performs a series of pair-wise alignments using the Myers and Miller global alignment algorithm (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line.

Parameters used in the comparison were:

Scoring matrix: Blosum62

First Gap: 12

Extending gap: 2

Results of the software analysis are shown in Table F for the global similarity and identity over the full length of the polypeptide sequences (excluding the partial polypeptide sequences). Percentage identity is given above the diagonal and percentage similarity is given below the diagonal.

The percentage identity between the full length SWI2/SNF2 polypeptide sequences of the SSO1653 subfamily, useful in performing the methods of the invention, ranges between 33 and 52% amino acid identity compared to SEQ ID NO: 30 (Table F).

The percentage identity between the ATPase domain of the SWI2/SNF2 polypeptide sequences of the SSO1653 subfamily, useful in performing the methods of the invention, ranges between 45 and 70% amino acid identity compared to the ATPase domain as
5 represented by SEQ ID NO: 111, comprised in SEQ ID NO: 30 (Table F1).

Table F: MatGAT results for global similarity and identity over the full length of the SWI2/SNF2 polypeptide sequences.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	
1. Synco_SNF2		48	38	33	52	46	48	38	33	37	37	37	38	36	47	34	40	49	37	41	41	41	41	36	38	37	42	40	42	43	43	42	43	42	43	42	48	48
2. Anava_SNF2	64	40	32	53	52	60	38	34	37	38	38	38	35	76	36	66	94	38	42	40	41	41	36	40	37	43	38	43	42	42	43	43	42	48	48	48		
3. Archaeon_RC-I_SNF2	57	60		34	39	40	40	41	34	40	42	42	39	36	41	36	32	41	38	36	36	37	37	36	39	38	38	33	37	37	36	37	37	39	39	39		
4. Bacce_ATCC10987_SNF2	49	48	52		33	34	33	33	33	32	34	34	31	34	32	32	26	32	34	30	30	28	28	33	32	35	29	27	29	30	30	29	29	33	33	34		
5. Crowa_SNF2	68	70	60	51		47	53	36	34	36	36	36	35	32	52	35	43	53	38	41	40	38	38	33	36	34	39	34	39	38	38	39	39	44	44	45		
6. Glovi_SNF2\	62	67	59	51	65		53	38	34	39	40	40	38	37	52	37	41	52	39	41	40	40	40	37	40	40	43	39	43	41	42	42	42	46	46	49		
7. Lyn_sp_SNF2	64	75	60	51	71	68		37	34	37	37	37	36	33	59	35	47	60	38	41	40	39	39	34	38	37	41	36	40	41	40	41	40	48	48	47		
8. Metac_C2A_SNF2	55	56	60	50	56	56	57		34	90	42	42	38	36	38	36	30	38	47	36	35	35	35	36	41	38	37	33	36	36	36	37	36	36	36	38		
9. Methu_JF-1_SNF2	53	53	55	48	56	52	53	52		34	35	35	32	33	33	31	27	34	33	30	31	31	31	33	33	32	30	29	31	32	32	31	32	31	33	33	34	
10. Metna_Goe1_SNF2	55	56	60	48	55	56	57	95	52		41	41	38	35	38	36	29	38	47	35	34	35	35	36	41	37	36	33	36	36	36	35	36	36	35	35	37	
11. Mycbo_SNF2	53	54	58	50	56	57	53	57	52	57		99	41	43	39	35	31	38	40	35	35	35	35	41	52	39	38	33	36	36	36	37	37	37	39	39	39	
12. Myctu_SNF2	53	54	58	50	56	57	53	57	52	57	99		41	42	39	35	31	38	40	35	35	35	35	41	52	39	38	33	36	36	36	37	37	37	39	39	39	
13. Myxxa_DK1622_SNF2	53	55	56	46	53	54	54	55	49	56	54	54		38	39	33	30	38	37	33	33	36	36	37	43	41	37	34	36	37	36	37	37	36	37	37	37	
14. Nocfa_IFM10152_SNF2	51	51	52	51	51	55	51	50	48	51	55	55	50		35	33	27	35	37	31	33	35	35	64	43	40	35	32	35	36	35	36	36	36	37	37	37	
15. Nodsp_SNF2	64	87	60	49	68	67	73	56	52	56	55	55	55	50		36	68	76	37	41	41	41	41	34	39	37	41	38	42	42	41	41	42	42	46	46	48	
16. Nos_sp_PCC7120_SNF2 II	53	56	58	51	56	55	55	56	51	56	54	54	52	51	55		29	37	37	33	31	30	30	32	35	32	32	29	32	32	32	31	31	34	34	35		
17. Nospu_PCC73102_SNF2	56	75	51	47	60	60	63	47	44	46	48	48	44	47	76	48		67	30	34	34	34	34	27	30	29	33	35	34	35	34	34	35	35	36	36	39	
18. Nostoc_SNF2	64	97	60	48	70	67	76	57	53	56	54	54	55	51	86	58	76		38	43	41	41	41	36	39	37	42	38	42	42	42	43	43	48	48	48		
19. Pelph_BU-1_SNF2	55	55	57	51	56	57	56	63	52	62	58	58	53	53	54	54	48	54		35	36	37	36	37	40	39	36	35	37	38	38	37	36	38	37	37	38	
20. Proma_CCMP1375_SNF2	58	60	56	47	60	58	62	56	51	55	52	52	50	48	59	52	51	59	52		63	60	60	32	34	36	58	57	61	62	61	62	61	61	41	41	40	
21. Proma_MIT9211_SNF2	58	58	55	46	60	58	61	55	50	54	53	53	50	50	59	52	51	59	54	78		66	66	32	35	35	61	61	66	66	65	65	66	65	42	42	40	

22. Proma_MIT9303_SNF2	58	59	54	45	59	57	59	54	50	54	51	50	52	49	58	49	50	58	51	76	80		99	35	38	37	73	75	83	82	80	84	83	82	44	44	40
23. Proma_MIT9313_SNF2	58	58	54	43	58	57	59	54	50	54	51	51	52	49	58	49	50	58	51	76	80	99		35	38	37	72	75	84	82	79	84	83	82	44	44	39
24. Rho_sp_RHA1_SNF2	51	51	51	52	52	54	51	52	49	52	55	55	49	75	50	50	48	51	54	49	51	49	49		43	40	36	31	35	35	35	35	35	37	37	38	
25. Saltr_CNB-440_SNF2	55	56	58	49	56	56	55	58	49	57	65	65	55	56	56	54	48	55	58	52	53	52	55		42	39	35	39	39	39	39	39	39	40	40	39	
26. Synth_IAM14863_SNF2	53	53	56	51	53	58	52	53	50	53	55	55	53	54	53	52	47	53	55	52	52	51	51	55	56		38	35	37	38	38	37	37	38	37	39	
27. Syn_sp_WH5701_SNF2	60	59	57	46	61	60	60	54	51	54	53	53	52	50	58	50	51	60	52	73	77	81	81	51	54	53		68	74	73	73	75	75	74	47	42	
28. Syn_sp_BL107_SNF2	56	56	53	44	57	57	57	50	47	50	49	49	48	47	55	48	53	56	51	73	75	83	83	48	50	51	79		78	85	93	78	79	85	42	38	
29. Syn_sp_CC9311_SNF2	59	60	56	44	60	60	61	55	51	54	52	52	51	49	59	52	51	60	52	77	81	89	89	49	54	51	83	86		84	83	89	91	85	45	41	
30. Syn_sp_CC9605_SNF2	59	60	57	46	60	59	61	55	52	55	52	52	52	50	59	52	51	60	54	78	81	88	88	51	54	53	82	90	91		90	85	85	92	45	45	41
31. Syn_sp_CC9902_SNF2	59	59	56	46	61	59	61	55	51	54	52	52	51	50	59	52	52	60	54	77	80	88	88	50	54	54	82	94	91	95		83	84	91	46	46	41
32. Syn_sp_RS9916_SNF2	59	60	56	45	59	59	60	56	50	55	53	53	52	50	58	52	51	60	53	79	81	90	90	49	55	51	83	87	94	92	92		89	85	46	46	41
33. Syn_sp_WH7805_SNF2	58	60	55	45	60	58	61	55	52	55	52	52	51	49	59	51	51	60	52	77	81	89	89	49	54	51	83	85	94	91	90	94		85	46	46	41
34. Syn_sp_WH8102_SNF2	60	60	56	45	62	59	61	54	51	55	53	53	51	50	59	51	51	60	54	78	81	89	89	51	54	53	83	91	92	96	96	92	92		46	46	41
35. Synel_PCC6301_SNF2	63	65	58	50	63	64	66	53	52	53	54	54	51	52	65	54	57	66	56	59	59	59	59	53	56	53	62	58	60	61	61	60	60	61		99	48
36. Synel_PCC7942_SNF2	63	65	58	51	63	64	66	53	52	53	54	54	51	52	65	53	57	66	56	59	59	59	59	53	56	53	62	58	60	61	61	60	60	61	99		48
37. Theel_BP-1_SNF2	60	62	56	51	63	65	63	55	51	53	55	55	51	52	61	54	55	63	54	57	55	54	54	53	54	56	58	55	56	56	57	56	56	64	64		

Table F1: MatGAT results for global similarity and identity between the ATPase domain of the SWI2/SNF2 polypeptide sequences.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1. ATPase_Synec_SNF2	65	52	50	70	63	63	63	63	54	50	52	52	52	52	51	65	48	45	65	53	54	55	57	57	52	52	57	49	56	57	58	58	56	57	63	63	62
2. ATPase_Anava_SNF2	77		55	50	70	69	69	72	54	50	53	54	54	54	52	85	51	60	97	53	55	52	55	55	54	51	56	47	56	56	55	56	56	56	65	65	67
3. ATPase_ArchaeonIRC-1_SNF2	70	74		51	53	56	56	56	54	50	53	56	55	53	54	56	52	37	54	55	49	49	52	52	53	54	52	44	51	52	52	51	51	53	53	55	
4. ATPase_Bacce_ATCC10987_SNF2	67	67	72		50	49	49	50	49	51	49	49	49	46	48	50	49	35	50	50	46	46	46	46	49	50	46	38	45	46	44	47	45	45	49	50	
5. ATPase_Crowa_SNF2	82	84	74	68		64	64	68	52	51	52	52	52	51	50	71	51	48	70	55	56	54	56	56	51	50	55	47	55	55	55	56	56	63	63	63	
6. ATPase_Glovi_SNF2	77	82	74	69	81		99	68	52	50	53	52	52	53	51	70	52	44	68	54	53	52	55	55	52	54	54	47	54	55	56	54	55	61	61	64	
7. ATPase_Glovi_SNF2\	77	82	74	69	81	99		68	52	50	53	52	52	53	51	70	52	44	68	54	53	52	55	55	52	54	54	47	54	55	56	54	55	61	61	64	
8. ATPase_Lyn_sp_SNF2	77	86	75	69	83	82	82		53	51	52	51	51	51	49	72	49	47	72	53	53	51	54	54	51	51	55	46	54	55	55	55	64	64	62	62	
9. ATPase_Metac_C2A_SNF2	70	71	74	67	71	71	71	72		49	92	55	55	51	52	55	53	36	53	65	49	49	51	51	53	53	51	43	51	51	52	51	51	51	50	54	54
10. ATPase_Methu_JF-1_SNF2	69	70	71	69	73	70	70	70	67		48	51	51	47	48	49	49	35	50	49	43	43	45	45	50	50	43	38	43	44	44	44	43	50	50	52	
11. ATPase_Metma_Goe1_SNF2	70	70	74	67	70	70	70	71	96	67		54	54	51	51	54	52	34	52	64	48	48	50	50	52	51	50	43	50	51	49	49	50	50	50	53	
12. ATPase_Mycbo_SNF2	68	70	73	67	70	71	71	69	69	69		99	54	60	54	50	36	54	55	47	46	48	48	59	52	50	40	48	48	48	48	48	49	51	51	54	
13. ATPase_Myctu_SNF2	68	70	73	67	70	71	71	69	69	69	68	99		54	60	54	49	36	54	55	47	46	48	48	59	52	50	40	48	48	48	48	49	51	51	54	
14. ATPase_Myxxa_DK1622_SNF2	67	69	70	63	67	70	70	69	69	69	63	69	68	68		55	53	46	35	53	50	45	47	49	49	52	56	49	41	48	49	49	48	51	51	52	
15. ATPase_Nocfa_IFM10152_SNF2	68	69	69	65	69	70	70	66	68	65	68	73	73	67		51	49	35	52	55	46	47	49	49	75	55	49	41	49	48	48	49	50	50	52	52	
16. ATPase_Nodsp_SNF2	77	91	76	69	85	82	82	85	71	70	71	71	71	69	70		52	58	86	53	55	53	57	57	53	52	56	47	56	57	56	56	65	65	68	68	
17. ATPase_Nos_sp_PCC7120_SNF2III	68	71	74	70	70	71	71	70	70	68	70	68	68	66	67	72		35	51	52	46	44	46	46	49	48	46	38	46	46	46	46	49	49	51	51	
18. ATPase_Nospu_PCC73102_SNF2	55	63	51	50	58	55	55	57	48	50	48	49	49	45	48	64	49		60	36	37	36	39	39	35	36	38	41	39	39	38	39	41	41	45	45	
19. ATPase_Nostoc_SNF2	77	99	74	67	84	82	82	85	71	69	70	70	70	69	69	92	71	63		53	55	53	55	55	53	51	55	46	56	56	55	55	56	65	65	66	66
20. ATPase_Pelph_BU-1_SNF2	70	71	72	70	73	73	73	71	79	68	79	72	72	68	71	72	70	49	71		51	52	55	55	55	56	52	45	54	54	53	54	52	52	54	54	
21. ATPase_Proma_CCMP1375_SNF2	71	71	69	66	73	72	72	71	67	64	67	66	66	64	63	73	64	49	72	68		71	71	71	48	50	69	61	70	70	70	71	70	57	57	56	56

22. ATPase_Proma_MIT9211_SNF2	72	70	69	63	73	72	72	72	67	63	67	67	66	63	65	72	65	48	71	69	83		74	73	47	49	69	61	72	73	72	71	72	72	56	56	54
23. ATPase_Proma_MIT9303_SNF2	74	73	70	64	75	75	72	69	65	69	66	66	66	64	65	74	64	51	72	69	84	87		99	50	53	85	75	87	88	86	87	86	88	59	59	57
24. ATPase_Proma_MIT9313_SNF2	74	73	69	64	75	75	72	69	65	69	66	66	66	64	65	74	64	51	72	69	84	87	99		50	53	85	75	87	88	86	87	86	88	59	59	57
25. ATPase_Rho_sp_RHA1_SNF2	69	71	70	66	72	70	69	71	67	71	74	74	74	67	83	73	68	50	71	73	66	65	66	66		55	50	42	50	50	50	50	50	52	52	52	
26. ATPase_Synth_IAM14863_SNF2	67	67	71	68	67	71	68	70	67	68	68	68	68	69	69	69	68	47	67	70	65	66	68	69			51	44	51	53	53	52	52	51	51	53	
27. ATPase_Syn_sp_WH5701_SNF2	74	73	69	64	75	73	73	68	65	68	67	67	67	65	65	73	64	51	73	70	83	85	93	93	66	67		73	84	84	84	85	85	86	59	59	57
28. ATPase_Syn_sp_BL107_SNF2	64	62	60	57	66	65	65	63	58	54	58	56	55	55	57	63	55	54	62	60	73	74	81	82	59	58	80		74	79	84	75	74	79	51	51	49
29. ATPase_Syn_sp_CC9311_SNF2	74	73	69	63	74	74	73	68	65	68	65	65	65	64	65	73	64	51	73	68	84	85	94	94	65	66	91	81		87	85	91	92	88	59	59	58
30. ATPase_Syn_sp_CC9605_SNF2	74	72	71	64	74	75	73	69	64	69	65	65	65	64	66	74	64	50	72	69	85	87	93	93	67	68	91	83	93		92	88	87	95	59	59	57
31. ATPase_Syn_sp_CC9902_SNF2	74	71	70	64	75	74	74	72	69	64	69	65	64	64	66	73	65	51	71	69	84	86	93	94	66	68	91	87	92	96		87	85	92	60	60	57
32. ATPase_Syn_sp_RS9916_SNF2	74	73	69	62	74	74	74	72	69	64	69	66	65	65	64	72	65	50	72	69	84	86	94	94	66	66	91	81	96	94	93		92	88	60	60	57
33. ATPase_Syn_sp_WH17805_SNF2	72	72	68	62	73	73	72	68	64	68	65	65	65	64	63	72	64	50	72	67	83	85	92	92	65	66	91	79	95	92	91	96		88	60	60	57
34. ATPase_Syn_sp_WH18102_SNF2	74	72	70	63	75	75	73	69	64	69	66	65	65	64	65	73	64	50	72	69	84	87	94	94	66	68	92	84	93	97	96	94	92		59	59	56
35. ATPase_Synel_PCC6301_SNF2	75	79	70	70	78	76	79	67	68	67	66	66	66	66	67	79	69	52	78	70	73	72	74	74	66	68	74	63	74	73	73	74	73	74		99	63
36. ATPase_Synel_PCC7942_SNF2	75	79	70	70	78	76	79	67	68	67	66	66	66	66	67	79	69	52	78	70	73	72	74	74	66	68	74	63	74	73	73	74	73	74	99		63
37. ATPase_Theel_BP-1_SNF2	75	78	72	69	79	79	76	69	71	69	68	68	66	66	67	79	70	54	78	71	71	70	71	71	69	69	72	63	72	71	71	72	71	76	76		76

Example 11: Identification of domains comprised in polypeptide sequences useful in performing the methods of the invention

The Integrated Resource of Protein Families, Domains and Sites (InterPro) database is an integrated interface for the commonly used signature databases for text- and sequence-based searches. The InterPro database combines these databases, which use different methodologies and varying degrees of biological information about well-characterized proteins to derive protein signatures. Collaborating databases include SWISS-PROT, PROSITE, TrEMBL, PRINTS, ProDom and Pfam, Smart and TIGRFAMs. Interpro is hosted at the European Bioinformatics Institute in the United Kingdom.

The relevant results of the InterPro scan of the polypeptide sequence as represented by SEQ ID NO: 30 are presented in Table G. SWI2/SNF2 polypeptides (or remodeling enzymes) share sequence similarity with helicases (particularly SF2 helicases), which are enzymes capable of catalyzing the separation of DNA strands using ATP hydrolysis. The sequence similarity is limited to the ATPase domain of both types of enzymes.

Table G: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 2.

InterPro accession number	InterPro decription	Originating database	Original accession number	Accession name
IPR000330	SNF2 related	Pfam	PF00176	SNF2_N
IPR001650	Helicase, C-terminal	Pfam	PF00271	Helicase_C
		SMART	SM00490	HELICc
		Profile	PS51194	Helicase_CTER
IPR014001	DEAD-like helicases, N-terminal	SMART	SM00487	DEXDc
IPR014021	Helicase superfamily a and 2 ATP binding	PROFILE	PS51192	Helicase_ATP_BIND_1

Example 12: Cloning of nucleic acid sequence as represented by SEQ ID NO: 29

Unless otherwise stated, recombinant DNA techniques are performed according to standard protocols described in (Sambrook (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York) or in Volumes 1 and 2 of Ausubel et al.

(1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfax (1993) by R.D.D. Croy, published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

5

The *Synechocystis* sp. PCC6803 SWI2/SNF2 gene was amplified by PCR using as template *Synechocystis* sp. PCC6803 genomic DNA. Primers prm08774 (SEQ ID NO: 113; sense, 5'-ggggacaagtttgtaaaaaagcaggcttaaacaatggcgactatccacggtaattgg-3') and prm08779 (SEQ ID NO: 114; reverse, complementary, 5'-ggggaccactttgtacaagaaagctgggttcaatcggacgcttcggctt - 3'), which include the AttB sites for Gateway recombination, were used for PCR amplification. PCR was performed using Hifi Taq DNA polymerase in standard conditions. A PCR fragment of the expected length (including attB sites) was amplified and purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined *in vivo* with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone". Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

10

15

Example 13: Expression vector construction using the nucleic acid sequence as represented by SEQ ID NO: 29

20

The entry clone comprising SEQ ID NO: 29 was subsequently used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice beta-expansin promoter (SEQ ID NO: 112) for expression in young expanding tissues was located upstream of this Gateway cassette.

25

After the LR recombination step, the resulting expression vector pExp::SWI2/SNF2 (Figure 8) was transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

30

Example 14: Plant transformation

See Example 5 above for rice transformation

35

Example 15: Phenotypic evaluation procedure15.1 Evaluation setup

Approximately 35 independent T0 rice transformants were generated. The primary transformants were transferred from a tissue culture chamber to a greenhouse for growing and harvest of T1 seed. Six events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes) and approximately 10 T1 seedlings lacking the transgene (nullizygotes) were selected by monitoring visual marker expression. The transgenic plants and the corresponding nullizygotes were grown side-by-side at random positions. Greenhouse conditions were of shorts days (12 hours light), 28°C in the light and 22°C in the dark, and a relative humidity of 70%.

Five T1 events were further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation but with more individuals per event. From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles.

Drought screen

Plants from five events (T2 seeds) were grown in potting soil under normal conditions until they approached the heading stage. They were then transferred to a "dry" section where irrigation was withheld. Humidity probes were inserted in randomly chosen pots to monitor the soil water content (SWC). When SWC went below certain thresholds, the plants were automatically re-watered continuously until a normal level was reached again. The plants were then re-transferred again to normal conditions. The rest of the cultivation (plant maturation, seed harvest) was the same as for plants not grown under abiotic stress conditions. Growth and yield parameters are recorded as detailed for growth under normal conditions.

Salt stress screen

The rice plants are grown on a substrate made of coco fibers and argex (3 to 1 ratio). A normal nutrient solution is used during the first two weeks after transplanting the plantlets in the greenhouse. After the first two weeks, 25 mM of salt (NaCl) is added to the nutrient solution comprising the components listed below.

- NPK Nutrient mix, 20-20-20 Peters professional (Scotts, Marysville, OH, USA) at a concentration of 1 kg/m³.
- Magnesium chelate, Chelal Mg (BMS, Bornem, Belgium) at 333.33 ml / m³
- Iron chelate, Libfer (CIBA, Bradford, UK) at 21.67 g / m³

- NaCl 1.425 kg / m³

Salt concentration is monitored on a weekly basis and additions are made where necessary. Plants are grown under these conditions until the start of grain filling. They are then transferred to a different compartment of the greenhouse where they are irrigated daily with fresh water until seed harvest. Growth and yield parameters are recorded as for growth under normal conditions.

Reduced nutrient (nitrogen) availability screen

The rice plants are grown in potting soil under normal conditions except for the nutrient solution. The pots are watered from transplantation to maturation with a specific nutrient solution containing reduced N nitrogen (N) content, usually between 7 to 8 times less. The rest of the cultivation (plant maturation, seed harvest) is the same as for plants not grown under abiotic stress. Growth and yield parameters are recorded as for growth under normal conditions.

15.2 Statistical analysis: F-test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F-test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F-test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F-test. A significant F-test value points to a gene effect, meaning that it is not only the mere presence or position of the gene that is causing the differences in phenotype.

15.3 Parameters measured

Biomass-related parameter measurement

From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles.

The plant aboveground area (or leafy biomass) was determined by counting the total number of pixels on the digital images from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from the different angles and was converted to a physical surface value expressed in square mm by calibration. Experiments show that the aboveground plant area measured this way correlates with the biomass of plant parts above ground. The above ground area is the area measured at

the time point at which the plant had reached its maximal leafy biomass. The early vigor is the plant (seedling) aboveground area three weeks post-germination.

To measure root-related parameters, plants were grown in specially designed pots with transparent bottoms to allow visualization of the roots. A digital camera recorded images through the bottom of the pot during plant growth. Increase in root biomass is expressed as an increase in total root biomass (measured as maximum biomass of roots observed during the lifespan of a plant); or as an increase in the root/shoot index (measured as the ratio between root mass and shoot mass in the period of active growth of root and shoot). Furthermore, the maximum biomass of roots above a certain thickness threshold observed during the lifespan of a plant is calculated (thick roots), as well as maximum biomass of roots below a certain thickness threshold (thin roots).

Seed-related parameter measurements

The mature primary panicles were harvested, counted, bagged, barcode-labelled and then dried for three days in an oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. The number of filled seeds was determined by counting the number of filled husks that remained after the separation step. The total seed weight per plant was measured by weighing all filled husks harvested from one plant. Total seed number per plant was measured by counting the number of husks harvested from a plant. Thousand Kernel Weight (TKW) is extrapolated from the number of filled seeds counted and their total weight. The Harvest Index (HI) in the present invention is defined as the ratio between the total seed weight per plant and the above ground area (mm²), multiplied by a factor 10⁶. The total number of flowers per panicle as defined in the present invention is the ratio between the total number of seeds and the number of mature primary panicles. The seed fill rate as defined in the present invention is the proportion (expressed as a %) of the number of filled seeds over the total number of seeds (or florets).

Example 16: Results of the phenotypic evaluation of the transgenic rice plants expressing the SWI2/SNF2 nucleic acid sequence, grown under normal conditions

The results of the evaluation of transgenic rice plants expressing the SWI2/SNF2 nucleic acid sequence, under normal growth conditions, are shown in Table H below.

There was an increase in the number of flowers per panicle, the total seed weight per plant, the total number of seeds, the number of filled seeds, and the harvest index of the transgenics compared to corresponding nullizygotes (controls).

- 5 **Table H** Results of the evaluation of transgenic rice plants expressing the SWI2/SNF2 nucleic acid sequence, under normal growth conditions.

	Average % increase of best performing events in T1 generation	Average % increase of best performing events in T2 generation
Number of flowers per panicle	11%	3%
Total seed weight per plant	13%	28%
Total number of seeds	14%	6%
Number of filled seeds	14%	25%
Harvest index	10%	25%

Example 17: Results of the phenotypic evaluation of the transgenic rice plants, grown under drought stress conditions

- 10 The results of the evaluation of transgenic rice plants expressing SWI2/SNF2 nucleic acid sequence, under drought stress growth conditions are presented in Table I.

There was an increase in the aboveground area, the total root biomass, the number of flowers per panicle, the seed fill rate, the total seed weight per plant, the total number of seeds, the number of filled seeds, and the harvest index of the transgenics compared to corresponding nullizygotes (controls).

Table I Results of the evaluation of transgenic rice plants expressing the SWI2/SNF2 nucleic acid sequence, under drought stress growth conditions.

	Average % increase of best performing events in T2 generation
Aboveground area	16%
Total root biomass	13%
Biomass thick roots	10%
Biomass thin roots	13%
Number of flowers per panicle	7%
Seed fill rate	28%
Total seed weight per plant	57%

Total number of seeds	44%
Number of filled seeds	54%
Harvest index	31%

Example 18: Examples of transformation of corn, alfalfa, cotton, soyabean, rapeseed/canola, wheat

See Example 5 above

Claims

1) A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an HpaG polypeptide comprising:

23) in increasing order of preference, at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the HpaG polypeptide sequence represented by SEQ ID NO: 2; and

24) an amino acid composition wherein the glycine content ranges between 13% and 25%, the glutamine content ranges between 13% and 20%, the cysteine content ranges between 0% and 1%, the histidine content ranges between 0% and 1%, and wherein tryptophan is absent.

2) Method according to claim 1, wherein said HpaG polypeptide further comprises one or more of the following motifs:

(i) (motif 1): G(G/E/D)(N/E)X(Q/R/P)Q(A/S)GX(N/D)G (SEQ ID NO: 3), wherein X on position 4 may be any amino acid, preferably one of S, N, P, R, or Q, and wherein X on position 9 may be any amino acid, preferably one of Q, E, S, or P; and

(ii) (motif 2): (P/A/V)S(P/Q/A)(F/L/Y)TQ(M/A)LM(H/N/Q)IV(G/M)(E/D/Q) (SEQ ID NO: 4),

3) Method according to claim 1 or 2, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding an HpaG polypeptide.

4) Method according to any preceding claim, wherein said nucleic acid encoding an HpaG polypeptide is represented by any one of the nucleic acids listed in Table A or a portion thereof, or a sequence capable of hybridising with any one of the nucleic acids given in Table A.

5) Method according to any preceding claim, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A.

6) Method according to any preceding claim, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

7) Method according to any one of claims 1 to 6, wherein said enhanced yield-related traits are obtained under non-stress conditions.

8) Method according to any one of claims 1 to 6, wherein said enhanced yield-related traits are obtained under abiotic stress conditions.

9) Method according to any one of claims 3 to 8, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

10) Method according to any one of claims 3 to 8, wherein said nucleic acid is operably linked to a green tissue-specific promoter, preferably to a protochlorophyllide reductase promoter, most preferably to a protochlorophyllide reductase promoter from rice.

11) Method according to any preceding claim, wherein said nucleic acid encoding an HpaG polypeptide is of prokaryotic origin, preferably from a plant pathogenic bacterium possessing a Type Three Secretion System (TTSS), further preferably from the family Pseudomonaceae, more preferably from the genus *Xanthomonas*, most preferably from *Xanthomonas axonopodis*.

12) Plant or part thereof, including seeds, obtainable by a method according to any preceding claim, wherein said plant or part thereof comprises a recombinant nucleic acid encoding an HpaG polypeptide.

13) Construct comprising:

(a) nucleic acid encoding an HpaG polypeptide as defined in claims 1 or 2;

(b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally

(c) a transcription termination sequence.

14) Construct according to claim 13, wherein said one of said control sequences is selected from:

(i) a constitutive promoter, preferably a GOS2 promoter, most preferably to a GOS2 promoter from rice; or

(ii) a green tissue-specific promoter, preferably a protochlorophyllide reductase promoter, most preferably a protochlorophyllide reductase promoter from rice.

15) Use of a construct according to claim 13 or 14 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

5 16) Plant, plant part or plant cell transformed with a construct according to any of claims 13 or 14.

17) Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:

- 10 (i) introducing and expressing in a plant a nucleic acid encoding an HpaG polypeptide as defined in claim 1 or 2; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

15 18) Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from increased expression of a nucleic acid encoding an HpaG polypeptide as defined in claim 1 or 2, or a transgenic plant cell derived from said transgenic plant.

20 19) Transgenic plant according to claim 12, 16 or 18, or a transgenic plant cell derived thereof, wherein said plant is a crop plant or a monocot or a cereal, such as rice, maize, wheat, barley, millet, rye, sorghum and oats.

25 20) Harvestable parts of a plant according to claim 19, wherein said harvestable parts are preferably seeds.

21) Products derived from a plant according to claim 19 and/or from harvestable parts of a plant according to claim 18.

30 22) Use of a nucleic acid encoding HpaG polypeptide in increasing yield, particularly in increasing seed yield, in plants relative to control plants.

23) A method for enhancing yield-related traits in plants relative to control plants, comprising increasing expression in a plant of a nucleic acid sequence encoding a SWITCH 2/ SUCROSE NON-FERMENTING 2 (SWI2/SNF2) polypeptide, which SWI2/SNF2 polypeptide comprises an ATPase domain comprising from N-terminus to C-terminus at least five, preferably six, more preferably seven, most preferably eight of the following motifs:

35

- (i) Motif I LADDMGLGK(T/S), as represented by SEQ ID N0: 103 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif I;
- 5 (ii) Motif Ia L(L/V/I)(V/I/L)(A/C)P(T/M/V)S(V/I/L)(V/I/L)XNW, as represented by SEQ ID N0: 104 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Ia;
- 10 (iii) Motif II DEAQ(N/A/H)(V/I/L)KN, as represented by SEQ ID N0: 105 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif II;
- (iv) Motif III A(L/M)TGTPXEN, as represented by SEQ ID N0: 106 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif III;
- 15 (v) Motif IV (L/I)XF(T/S)Q(F/Y), as represented by SEQ ID N0: 107 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif IV;
- (vi) Motif V S(L/V)KAGG(V/T/L)G(L/I)(N/T)LTXA(N/S/T)HV, as represented by SEQ ID N0: 108 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif V;
- 20 (vii) Motif Va DRWWNPAVE, as represented by SEQ ID N0: 109 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif Va;
- 25 and
- (viii) Motif VI QA(T/S)DR(A/T/V)(F/Y)R(I/L)GQ, as represented by SEQ ID N0: 110 or a motif having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the sequence of Motif VI,
- 30 where X in Motif Ia, Motif III, Motif IV, and Motif V, is any amino acid.

24) Method according to claim 23, wherein said SWI2/SNF2 polypeptide, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, tends to cluster with the SSO1653 clade of SWI2/SNF2 polypeptides comprising the polypeptide sequence as represented by SEQ ID NO: 30 rather than with any other SWI2/SNF2 clade.

35

25) Method according to claim 23 or 24, wherein said SWI2/SNF2 polypeptide comprises an ATPase domain having in increasing order of preference at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the ATPase domain as represented by SEQ ID NO: 111, comprised in SEQ ID NO: 30.

5

26) Method according to any one of claims 23 to 25, wherein said SWI2/SNF2 polypeptide has in increasing order of preference at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more sequence identity to the SWI2/SNF2 polypeptide as represented by SEQ ID NO: 30 or to any of the polypeptide sequences given in Table E herein.

10

27) Method according to any one of claims 23 to 26, wherein said nucleic acid sequence encoding a SWI2/SNF2 polypeptide is represented by any one of the nucleic acid sequence SEQ ID NOs given in Table E or a portion thereof, or a sequence capable of hybridising with any one of the nucleic acid sequences SEQ ID NOs given in Table E.

15

28) Method according to any one of claims 23 to 27, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the SEQ ID NOs given in Table E.

29) Method according to any one of claims 23 to 28, wherein said increased expression is effected by introducing and expressing in a plant a nucleic acid sequence encoding a SWI2/SNF2 polypeptide.

20

30) Method according to any one of claims 23 to 29, wherein said yield-related traits are one or more of: (i) increased number of flowers per panicle; (ii) increased total seed weight per plant; (iii) increased number of (filled) seeds; or (iv) increased harvest index.

25

31) Method according to any one of claims 23 to 30, wherein said yield-related traits are enhanced in plants grown under abiotic stress conditions, preferably under water stress conditions, most preferably under drought stress conditions, relative to control plants grown under comparable stress conditions.

30

32) Method according to claim 31, wherein said enhanced yield-related traits are one or more of: (i) increased aboveground area; (ii) increased total root biomass; (iii) increased thick root biomass; (iv) increased thin root biomass; (v) increased number of flowers per panicle; (vi) increased seed fill rate; (vii) increased total seed weight per plant; (viii) increased number of (filled) seeds; or (ix) increased harvest index.

35

33) Method according to any one of claims 23 to 32, wherein said nucleic acid sequence is operably linked to a tissue-specific promoter, preferably to a promoter capable of preferentially expressing the nucleic acid sequence in young expanding tissues, most preferably to a beta-expansin promoter.

34) Method according to any one of claims 23 to 33, wherein said nucleic acid sequence encoding a SWI2/SNF2 polypeptide is from a microbial genome, further preferably from archaea or bacteria, more preferably from cyanobacteria, such as *Synechocystis* sp., *Nostoc* sp., *Synechococcus* sp., *Prochlorococcus* sp., *Anaebena* sp., *Gloeobacter* sp., or *Thermosynechococcus* sp., more preferably from *Synechocystis* sp., most preferably from *Synechocystis* sp. PCC6803.

35) Plants, parts thereof (including seeds), or plant cells obtainable by a method according to any one of claims 23 to 34, wherein said plant, part or cell thereof comprises an isolated nucleic acid transgene encoding a SWI2/SNF2 polypeptide.

36) Construct comprising:

- (a) A nucleic acid sequence encoding a SWI2/SNF2 polypeptide as defined in any one of claims 23 to 28;
- (b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
- (c) a transcription termination sequence.

37) Construct according to claim 36, wherein said one of said control sequences is a tissue-specific promoter, preferably a promoter for expression in young expanding tissues, most preferably a beta-expansin promoter.

38) Use of a construct according to claims 36 or 37 in a method for making plants having enhanced yield-related traits relative to control plants.

39) Plant, plant part or plant cell transformed with a construct according to claim 36 or 37.

40) Method for the production of transgenic plants having enhanced yield-related traits relative to control plants, comprising:

- (i) introducing and expressing in a plant a nucleic acid sequence encoding a SWI2/SNF2 polypeptide as defined in any one of claims 23 to 28; and

- (ii) cultivating the plant cell under conditions promoting plant growth and development.

41) Transgenic plant having enhanced yield-related traits relative to control plants, resulting from increased expression of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide as defined in any one of claims 23 to 28, or a transgenic plant cell derived from said transgenic plant.

42) Transgenic plant according to claim 35, 39 or 41, wherein said plant is a crop plant or a monocot or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum and oats, or a transgenic plant cell derived from said transgenic plant.

43) Harvestable parts of a plant according to claim 42, wherein said harvestable parts are preferably seeds.

44) Products derived from a plant according to claim 42 and/or from harvestable parts of a plant according to claim 43.

45) Use of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide as defined in any one of claims 23 to 28 in enhancing yield-related traits in plants, preferably in increasing one or more of: (i) increased number of flowers per panicle; (ii) increased total seed weight per plant; (iii) increased number of (filled) seeds; or (iv) increased harvest index.

46) Use of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide as defined in any one of claims 23 to 28 in enhancing yield-related traits in plants, wherein said yield-related traits are enhanced in plants grown under abiotic stress conditions, preferably under water stress conditions, most preferably under drought stress conditions, relative to control plants grown under comparable stress conditions.

47) Use of a nucleic acid sequence according to claim 45, wherein said enhanced yield-related traits are one or more of: (i) increased aboveground area; (ii) increased total root biomass; (iii) increased thick root biomass; (iv) increased thin root biomass; (v) increased number of flowers per panicle; (vi) increased seed fill rate; (vii) increased total seed weight per plant; (viii) increased number of (filled) seeds; or (ix) increased harvest index.

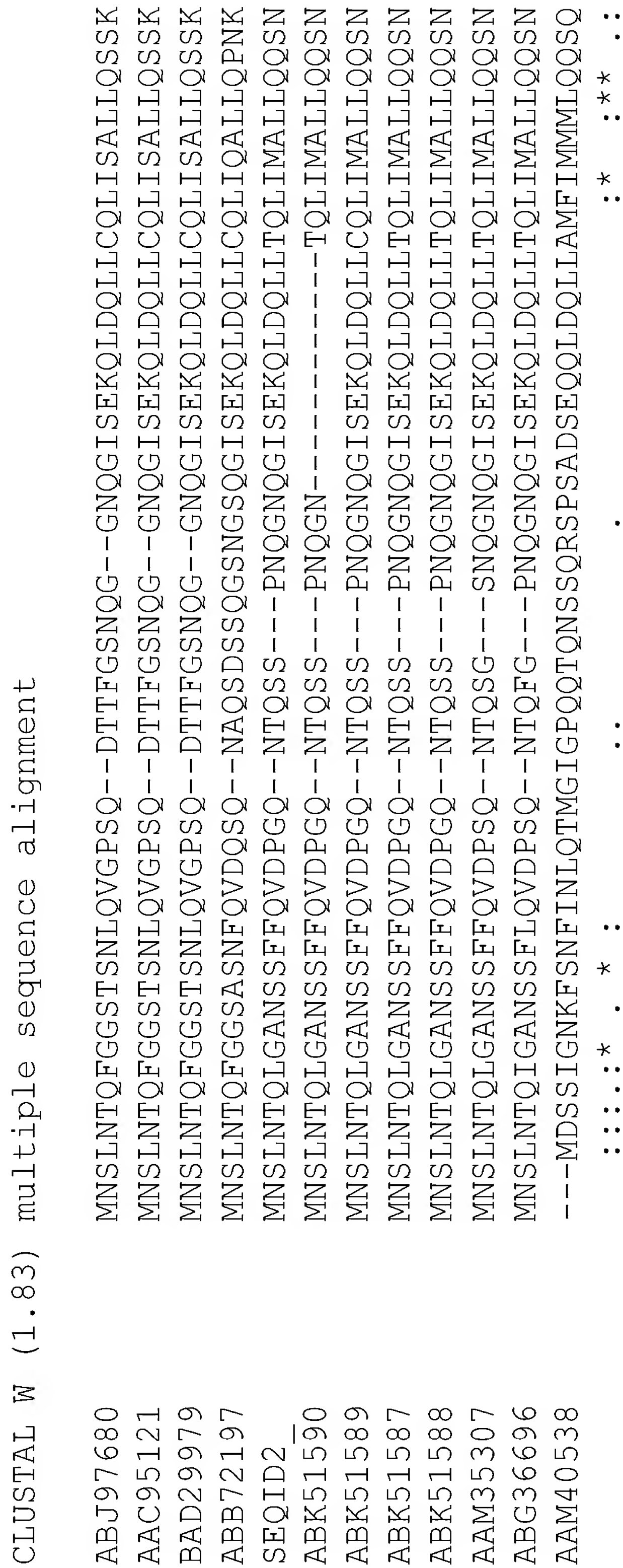


FIGURE 1

ABJ97680	NAEEGKG-QGGDNGGQGGNSQQAGQQNG-PSPFTQMLMHI VGEILQAQNGGAGGGGFG
AAC95121	NAEEGKG-QGGDNGGQGGNSQQAGQQNG-PSPFTQMLMHI VGEILQAQNGGAGGGGFG
BAD29979	NAEEGKG-QGGDNGGQGGNSQQAGQQNG-PSPFTQMLMHI VGEILQAQNGGAGGGGFG
ABB72197	NAEEGKG-QQG-----GENNQAGKENG-ASPLTQMLMNI VGEILQAQNAAGSSGGDFG
SEQID2_	NAEQGGQGGQGGDSGGQ GGNPRQAGQNSGSPSQYTQALMNI VGDI ILQAQNGGFGGGFGG
ABK51590	NAEQGGQGGQGGDSGGQGGNPRQAGQNSGSPSQYTQALMNI VGDILQAQNGGFGGGFGG
ABK51589	NAEQGGQGGQGGDSGGQGGNPRQAGQNSGSPSQYTQALMNI VGDILQAQNGGFGGGFGG
ABK51587	NAEQGGQGGQGGDSGGQGGNPRQAGQNSGSPSQYTQALMNI VGD-----GFGGGFGG
ABK51588	NAEQGGQGGQGGDSGGQGGNPRQAGQNSGSPSQYTQALMNI VGDILQAQN-----
AAM35307	NAEQGGQGGQGGDSGGQGGNRQAGQNSGSPSQYTQMLMNI VGDILQAQNGGFGGGFGG
ABG36696	NADQ----GQGGDSGGQGGNSRQAGQPNGSPSAYTQMLMNI VGDILQAQNGGFGGGFGG
AAM40538	GSDADQE-----CGDEQPQSGQQDG-VSPLTQMLMQI VMQLMQNQGGAGMGGTSLG
	.::
	* ; *:*: ;* * * * * *

FIGURE 1 (continued)

FIGURE 1 (continued)

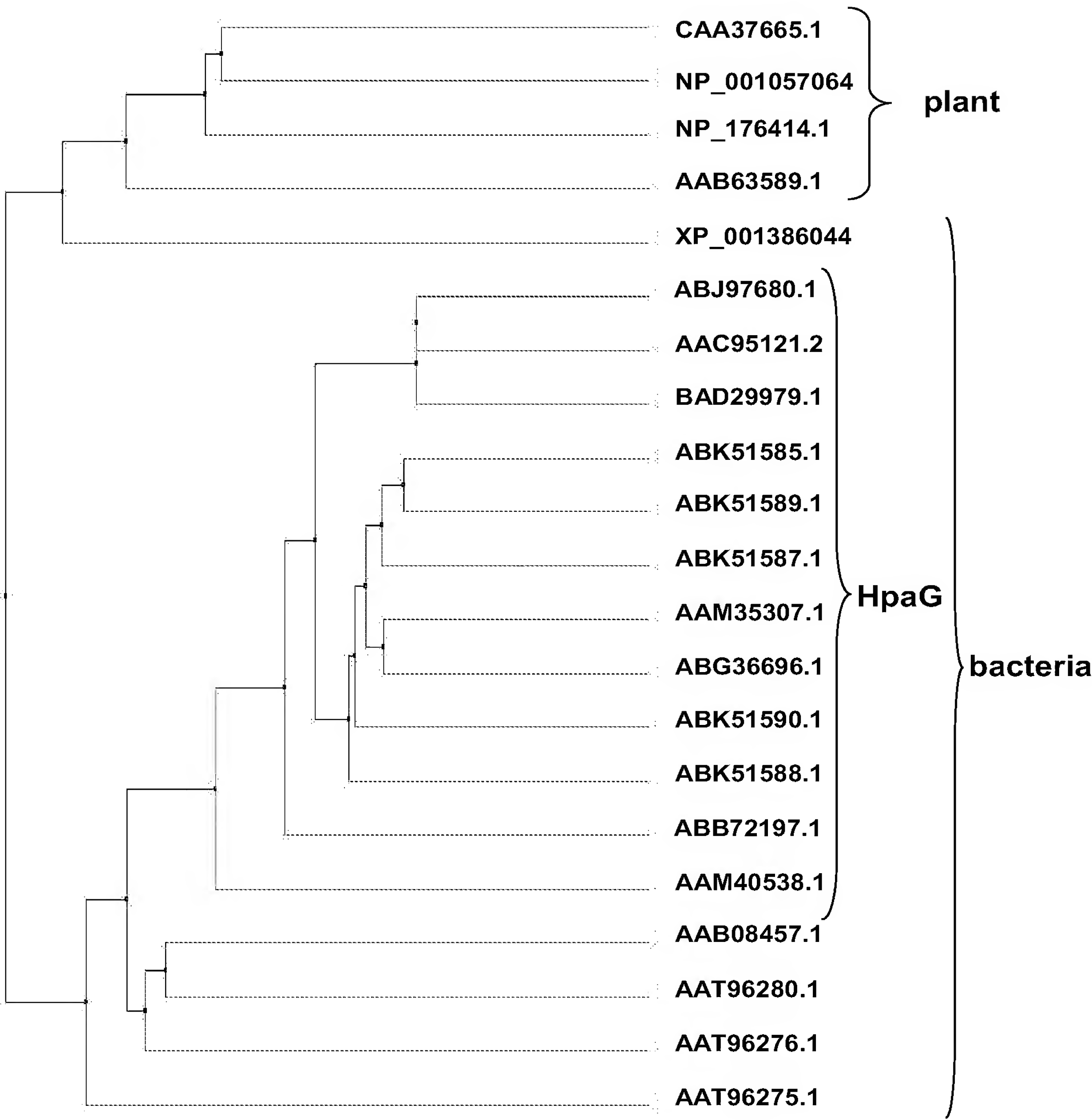


FIGURE 2

5/96

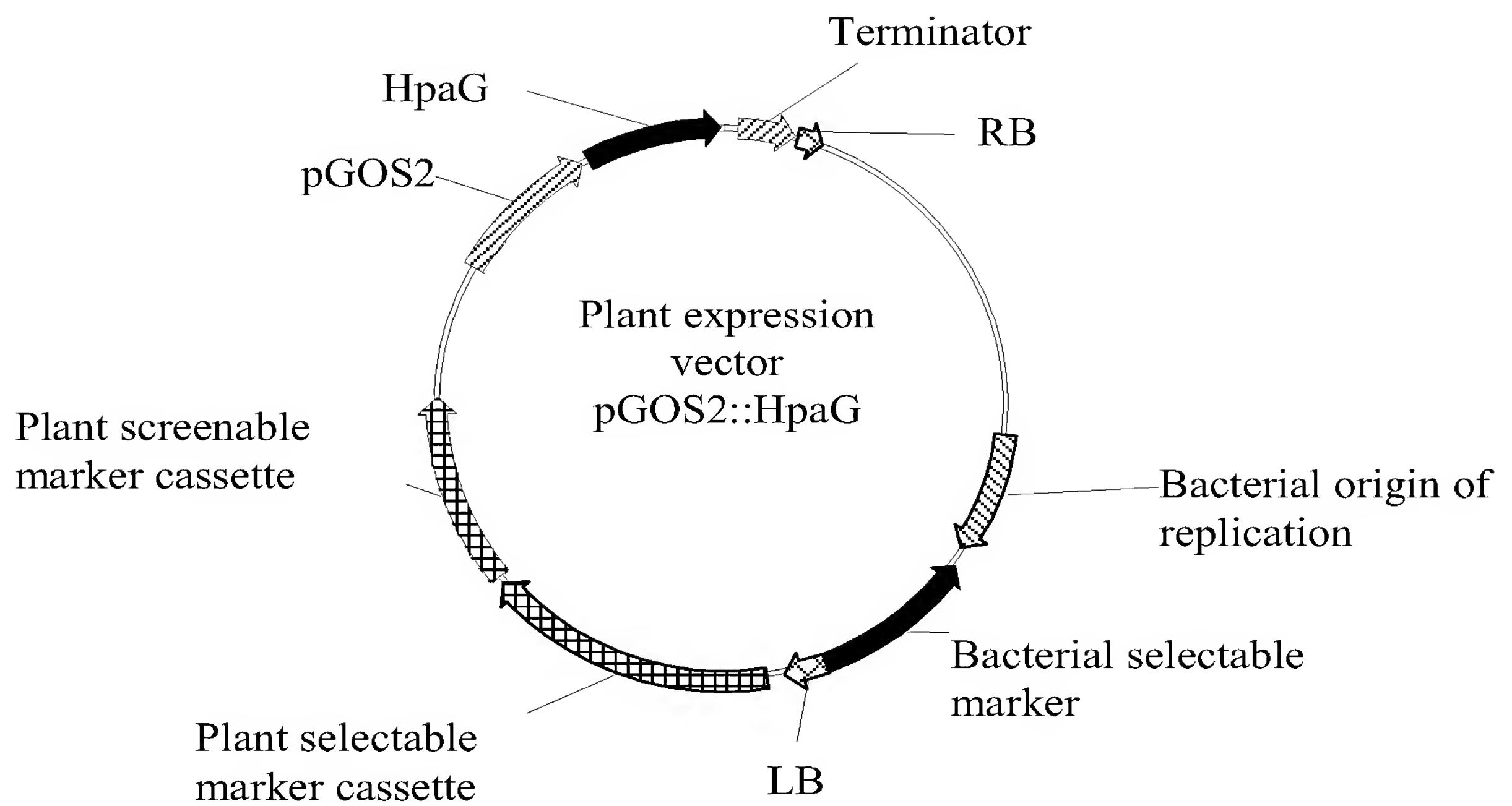


FIGURE 3

6/96

SEQ ID NO: 1, EF050509.1, *Xanthomonas axonopodis* elicitor of hypersensitive response HpaG (hpaG) gene, complete cds

ATGAATTCTTTGAACACACAGCTCGGGCGCCAACTCGTCCTTCTTTCAGGTTGACCCCGGCCAGAAC
ACGCAATCTAGTCCGAACCAGGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGCTGCTGACC
CAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAATGCCGAGCAGGGTCAGGGTCAAGGCCAGGGT
GGTGA CTCTGGCGGTCAGGGCGGCAATCCGCGGCAGGCCGGGCAGTCCAACGGCTCCCCCTCGCAA
TACACCCAGGCGCTGATGAATATCGTCGGAGACATTCTCCAGGCGCAGAATGGTGGCGGCTTCGGC
GGCGGCTTTGGTGGTGGCTTCGGTGGCATCCTCGTCACCAGCCTTGCGAGCGACACCGGATCGATG
CAGTAA

SEQ ID NO: 2, ABK51582.1, elicitor of hypersensitive response HpaG [*Xanthomonas axonopodis*]

MNSLNTQLGANSSFFQVDPGQNTQSSPNQGNQGI SEKQLDQLLTQLIMALLQQSNNAEQGGQGQGQ
GDSGGQGGNPRQAGQSNGSPSQYTQALMNIVGDILQAQNGGGFGGGFGGGFGGGILVTSLASDTGSM
Q

SEQ ID NO: 3, conserved motif 1

G (G/E/D) (N/E) X (Q/R/P) Q (A/S) GX (N/D) G

SEQ ID NO: 4, conserved motif 2

(P/A/V) S (P/Q/A) (F/L/Y) TQ (M/A) LM (H/N/Q) IV (G/M) (E/D/Q)

SEQ ID NO: 5, constitutive promoter GOS2

AATCCGAAAAGTTTCTGCACCGTTTTTCACCCCCTAACTAACAATATAGGGAACGTGTGCTAAATAT
AAAATGAGACCTTATATATGTAGCGCTGATAACTAGAACTATGCAAGAAAAACTCATCCACCTACT
TTAGTGGCAATCGGGCTAAATAAAAAAGAGTCGCTACACTAGTTTCGTTTTCTTAGTAATTAAGT
GGGAAAATGAAATCATTATTGCTTAGAATATACGTTACATCTCTGT CATGAAGTTAAATTATTCG
AGGTAGCCATAATTGTCATCAAACCTCTTCTTGAATAAAAAAATCTTTCTAGCTGAACTCAATGGGT
AAAGAGAGAGATTTTTTTTTTAAAAAATAGAATGAAGATATTCTGAACGTATTGGCAAAGATTTAAA
CATATAATTATATAATTTTATAGTTTGTGCATTCGTCATATCGCACATCATTAAGGACATGTCTTA
CTCCATCCCAATTTTTTATTTAGTAATTAAAGACAATTGACTTATTTTTTATTATTTATCTTTTTTCG
ATTAGATGCAAGGTACTTACGCACACACTTTGTGCTCATGTGCATGTGTGAGTGCACCTCCTCAAT
ACACGTTCAACTAGCAACACATCTCTAATATCACTCGCCTATTTAATACATTTAGGTAGCAATATC
TGAATTCAAGCACTCCACCATCACCAGACCACTTTTAATAATATCTAAAATACAAAAAATAATTTT
ACAGAATAGCATGAAAAGTATGAAACGAACTATTTAGGTTTTTTCACATACAAAAAAGAAATT
TTGCTCGTGC GCGAGCGCCAATCTCCCATATTGGGCACACAGGCAACAACAGAGTGGCTGCCCA
GAACAACCCACAAAAAACGATGATCTAACGGAGGACAGCAAGTCCGCAACAACCTTTTAACAGCAG
GCTTTGCGGCCAGGAGAGAGGAGGAGAGGCAAAGAAAACCAAGCATCCTCCTTCTCCCATCTATAA
ATTCTTCCCCCTTTTCCCCTCTCTATATAGGAGGCATCCAAGCCAAGAAGAGGGAGAGCACCAAG
GACACGCGACTAGCAGAAGCCGAGCGACCGCCTTCTCGATCCATATCTTCCGGTCGAGTTCTTGGT
CGATCTCTTCCCTCCTCCACCTCCTCCTCACAGGGTATGTGCCTCCCTTCGGTTGTTCTTGGATTT
ATTGTTCTAGGTTGTGTAGTACGGGCGTTGATGTTAGGAAAGGGGATCTGTATCTGTGATGATTCC
TGTTCTTGGATTTGGGATAGAGGGGTCTTGATGTTGCATGTTATCGGTTCCGGTTTGATTAGTAGT
ATGGTTTTTCAATCGTCTGGAGAGCTCTATGGAAATGAAATGGTTTAGGGATCGGAATCTTGCGATT
TTGTGAGTACCTTTTGTGTTGAGGTAAAATCAGAGCACCGGTGATTTTGCTTGGTGTAATAAAGTAC
GGTTGTTTGGTCCTCGATTCTGGTAGTGCTTCTCGATTTGACGAAGCTATCCTTTGTTTATTC
CCTATTGAACAAAAATAATCCAACCTTTGAAGACGGTCCCGTTGATGAGATTGAATGATTGATTCTT
AAGCCTGTCCAAAATTTTCGCAGCTGGCTTGTTTAGATACAGTAGTCCCCATCACGAAATTCATGGA

FIGURE 4

7/96

AACAGTTATAATCCTCAGGAACAGGGGATTCCCTGTTCTTCCGATTTGCTTTAGTCCCAGAATTTT
TTTTCCCAAATATCTTAAAAAGTCACTTTCTGGTTCAGTTCAATGAATTGATTGCTACAAATAATG
CTTTTATAGCGTTATCCTAGCTGTAGTTCAGTTAATAGGTAATACCCCTATAGTTTGTAGTCAGGAGA
AGAACTTATCCGATTTCTGATCTCCATTTTAAATTATATGAAATGAACTGTAGCATAAGCAGTATT
CATTTGGATTATTTTTTTTTTATTAGCTCTCACCCCTTCATTATTCTGAGCTGAAAGTCTGGCATGAA
CTGTCCTCAATTTTGTTCATTCACATCGATTATCTATGCATTATCCTCTTGTATCTACCTGT
AGAAGTTTCTTTTGGTTATTCCTTGACTGCTTGATTACAGAAAGAAATTTATGAAGCTGTAATCG
GGATAGTTATACTGCTTGTCTTATGATTCATTTCTTTGTGCAGTTCTTGGTGTAGCTTGCCACT
TTCACCAGCAAAGTTC

SEQ ID NO: 6, green tissue specific promoter PCR

TTGCAGTTGTGACCAAGTAAGCTGAGCATGCCCTTAACTTCACCTAGAAAAAAGTATACTTGGCTT
AACTGCTAGTAAGACATTTTCAAGACTGAGACTGGTGTACGCATTTTCATGCAAGCCATTACCACTTT
ACCTGACATTTTGGACAGAGATTAGAAATAGTTTCGTACTACCTGCAAGTTGCAACTTGAAAAGTG
AAATTTGTTTCTTGTCTAATATATTGGCGTGTAATTCCTTTTATGCGTTAGCGTAAAAAGTTGAAATT
TGGGTCAAGTTACTGGTCAGATTAACCAGTAACTGGTTAAAGTTGAAAGATGGTCTTTTAGTAATG
GAGGGAGTACTACACTATCCTCAGCTGATTTAAATCTTATTCGTCGGTGGTGAATTCGTCAATCT
CCCAACTTAGTTTTTCAATATATTCATAGGATAGAGTGTGCATATGTGTGTTTATAGGGATGAGTC
TACGCGCCTTATGAACACCTACTTTTGTACTGTATTTGTCAATGAAAAGAAAATCTTACCAATGCT
GCGATGCTGACACCAAGAAGAGGCGATGAAAAGTGCAACGGATATCGTGCCACGTCGGTTGCCAAG
TCAGCACAGACCCAATGGGCCTTTCCTACGTGTCTCGGCCACAGCCAGTCGTTTACCGCACGTTCA
CATGGGCACGAACCTCGCGTCATCTTCCACGCAAAACGACAGATCTGCCCTATCTGGTCCCACCCA
TCAGTGGCCACACCTCCCATGCTGCATTATTTGCGACTCCCATCCCGTCCTCCACGCCCAAACAC
CGCACACGGGTCGCGATAGCCACGACCCAATCACACAACGCCACGTCACCATATGTTACGGGCAGC
CATGCGCAGAAGATCCCGCGACGTGCTGTCCCCCGTGTGCGTTACGAAAAAATATCCCACCACGT
GTCGCTTTTACAGGACAATATCTCGAAGGAAAAAATCGTAGCGGAAATCCGAGGCACGAGCTGC
GATTGGCTGGGAGGCGTCCAGCGTGGTGGGGGGCCACCCCTTATCCTTAGCCCGTGGCGCTCCT
CGCTCCTCGGGTCCGTGTATAAATACCCTCCGGAACCTCACTCTTGCTGGTCACCAACACGAAGCAA
AAGGACACCAGAAACATAGTACACTTGAGCTCACTCCAACTCAAACACTCACACCA

SEQ ID NO: 7, EF042294, Synthetic construct mutant elicitor of hypersensitive response HpaG_T44C gene, complete cds

ATGAATTCTTTGAACACACAGCTCGGCGCCAACTCGTCCTTCTTTTCAAGTTGACCCCGGCCAGAAC
ACGCAATCTAGTCCGAACCAGGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGCTGCTGTGC
CAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAATGCCGAGCAGGGTCAGGGTCAAGGCCAGGGT
GGTGAATCTTGGCGGTGAGGGCGGCAATCCGCGGCAGGCCGGGCGAGTCCAACGGCTCCCCCTCGCAA
TACACCCAGGCGCTGATGAATATCGTCGGAGACATTCTCCAGGCGCAGAATGGTGGCGGCTTCGGC
GGCGGCTTTGGTGGTGGCTTCGGTGGCATCCTCGTCACCAGCCTTGCGAGCGACACCGGATCGATG
CAGTAA

SEQ ID NO: 8, ABK51589, mutant elicitor of hypersensitive response HpaG_T44C [synthetic construct]

MNSLNTQLGANSSFFQVDPGQNTQSSPNQGNQGI SEKQLDQLLQQLIMALLQQSNNAEQGQGQGQG
GDSGGQGGNPRQAGQSNGSPSQYTQALMNIVGDILQAQNGGGFGGGFGGGFGGILVTSLASDTGSM
Q

FIGURE 4 (continued)

8/96

SEQ ID NO: 9, EF042292, Synthetic construct mutant elicitor of hypersensitive response HpaG-T gene, complete cds

ATGAATTCTTTGAACACACAGCTCGGCGCCAACTCGTCCTTCTTTTCAGGTTGACCCCGGCCAGAAC
ACGCAATCTAGTCCGAACCAGGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGCTGCTGACC
CAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAATGCCGAGCAGGGTCAGGGTCAAGGCCAGGGT
GGTGA CTCTGGCGGT CAGGGCGGCAATCCGCGGCAGGCCGGGCAGTCCAACGGCTCCCCCTCGCAA
TACACCCAGGCGCTGATGAATATCGTCGGAGACGGCTTCGGCGGCGGCTTTGGTGGTGGCTTCGGT
GGCATCCTCGTCACCAGCCTTGCGAGCGACACCGGATCGATGCAGTAA

SEQ ID NO: 10, ABK51587, mutant elicitor of hypersensitive response HpaG-T [synthetic construct]

MNSLNTQLGANSSFFQVDPGQNTQSSPNQGNQGI SEKQLDQLLTQLIMALLQQSNNAEQGQGQGQG
GDSGGQGGNPRQAGQSNNGSPSQYTQALMNIVGDGFGGGFGGGFGGGILVTSLASDTGSMQ

SEQ ID NO: 11, 21106495:2613-3026 Xanthomonas axonopodis pv. citri str. 306, section 45 of 469 of the complete genome

TTACTGCATCGATCCGGTGTCTGCTCGCAAGGCTGGTGCCGAGGCTGGTGCCGAGGCCGCCGCCGAA
GCCACCACCAAAGCCGCCGCCGAAGCCACCACCATTCTGCGCCTGGAGAATGTCTCCGACGATATT
CATCAGCATCTGGGTGTATTGCGAGGGGGAGCCGTTGGACTGACCGGCCTGCTGCCGATTGCCGCC
CTGACCACCAGAGTCACCACCCTGGCCTTGACCCTGACCCTGCTCGGCATTGTTGCTCTGCTGAAG
CAGGGCCATGATGAGCTGGGTCAGCAGCTGGTCCAGTTGCTTTTCCGAGATGCCCTGGTTGCCCTG
GTTCAAGCAGATTGCGTGTTCTGGCTGGGGTCAACCTGAAAGAAGGACGAGTTGGCGCCGAGCTG
TGTGTTCAAAGAATTCAT

SEQ ID NO: 12, AAM35307, Hpa1 protein [Xanthomonas axonopodis pv. citri str. 306]

MNSLNTQLGANSSFFQVDPNQNTQSGSNQGNQGI SEKQLDQLLTQLIMALLQQSNNAEQGQGQGQG
GDSGGQGGNRQQAGQSNNGSPSQYTQMLMNIVGDILQAQNGGGFGGGFGGGFGGGLGTSLGTSLASD
TGSMQ

SEQ ID NO: 13, EF042295, Synthetic construct mutant elicitor of hypersensitive response HpaG-N gene, complete cds

ATGAATTCTTTGAACACACAGCTCGGCGCCAACTCGTCCTTCTTTTCAGGTTGACCCCGGCCAGAAC
ACGCAATCTAGTCCGAACCAGGGCAACACCCAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAAT
GCCGAGCAGGGTCAGGGTCAAGGCCAGGGTGGTGA CTCTGGCGGT CAGGGCGGCAATCCGCGGCAG
GCCGGGCAGTCCAACGGCTCCCCCTCGCAATACACCCAGGCGCTGATGAATATCGTCGGAGACATT
CTCCAGGCGCAGAATGGTGGCGGCTTCGGCGGCGGCTTTGGTGGTGGCTTCGGTGGCATCCTCGTC
ACCAGCCTTGCGAGCGACACCGGATCGATGCAGTAA

SEQ ID NO: 14, ABK51590, mutant elicitor of hypersensitive response HpaG-N [synthetic construct]

MNSLNTQLGANSSFFQVDPGQNTQSSPNQGNQNTQLIMALLQQSNNAEQGQGQGQGQGGDSGGQGGNPRQ
AGQSNNGSPSQYTQALMNIVGDILQAQNGGGFGGGFGGGFGGGILVTSLASDTGSMQ

SEQ ID NO: 15, EF042293, Xanthomonas axonopodis HpaG_G gene, complete cds

ATGAATTCTTTGAACACACAGCTCGGCGCCAACTCGTCCTTCTTTTCAGGTTGACCCCGGCCAGAAC
ACGCAATCTAGTCCGAACCAGGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGCTGCTGACC

FIGURE 4 (continued)

9/96

CAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAATGCCGAGCAGGGTCAGGGTCAAGGCCAGGGT
GGTGA CTCTGGCGGT CAGGGCGGCAATCCGCGGCAGGCCGGGCAGTCCAACGGCTCCCCCTCGCAA
TACACCCAGGCGCTGATGAATATCGTCGGAGACATTCTCCAGGCGCAGAATGGCTTTATCCTCGTC
ACCAGCCTTGCGAGCGACACCGGATCGATGCAGTAA

SEQ ID NO: 16, ABK51588, HpaG_G [Xanthomonas axonopodis]

MNSLNTQLGANSSFFQVDPGQNTQSSPNQGNQGI SEKQLDQLLTQLIMALLQQSNNAEQGQGQGQG
GDSGGQGGNPRQAGQSNGPSQYTQALMNIVGDILQAQNGFILVTSLASDTGSMQ

SEQ ID NO: 17, DQ643828, Xanthomonas smithii subsp. smithii Hrp gene, complete cds

ATGAATTCTTTGAACACACAGATCGGCGCCAACTCGTCCTTCTTGCAGGTGACCCGAGCCAGAAC
ACGCAATTCGGTCCGAACCAGGGCAATCAAGGCATCTCGGAAAAGCAGCTGGACCAGCTGCTGACC
CAGCTCATCATGGCCCTGCTTCAGCAGAGCAACAATGCCGACCAGGGTCAGGGTGGTGA CTCTGGT
GGTCAAGGCGGCAATTCGCGGCAGGCCGGGCAGCCCAATGGTTCCCCCTCGGCATACACCCAGATG
CTGATGAATATCGTCGGAGACATTCTCCAGGCGCAGAATGGTGGTGGCTTCGGCGGCGGGTTCGGC
GGTGGCTTTGGTGGCGGGCTCGGCACCAGCCTCGGCAGCAGCCTTGCGAGCGACACCGGATCGATG
CAGTAA

SEQ ID NO: 18, ABG36696, Hrp [Xanthomonas smithii subsp. smithii]

MNSLNTQIGANSSFLQVDPSQNTQFGPNQGNQGI SEKQLDQLLTQLIMALLQQSNNADQGQGGDSG
GQGGNSRQAGQPNGSPSAYTQMLMNIVGDILQAQNGGGFGGGFGGGFGGGLGTSLGSSLASDTGSM
Q

SEQ ID NO: 19, gi|116292746:1016-1435 Xanthomonas oryzae pv. oryzae strain JXOIII hrp gene cluster, partial sequence

ATGAACTCTTTGAACACACAATTCGGCGGCAGCACGTCCAACCTTCAGGTTGGCCCAAGCCAGGAC
ACAACGTTTCGGTTCGAACCAGGGCGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGTTGCTG
TGCCAGCTCATCTCGGCCCTGCTTCAGTCGAGCAAAAATGCTGAGGAGGGTAAGGGTCAGGGTGGC
GATAATGGCGGTGGCCAGGGCGGCAATTCGCAGCAGGCCGGGCAGCAGAATGGCCCCCTCGCCATTC
ACCCAGATGCTGATGCATATCGTCGGAGAGATTCTCCAGGCGCAGAATGGTGGTGGTGGTGGTGGC
GGCGGTTTCGGCGGCGGGTTCGGCGGCGACTTTAGTGGCGACCTCGGCCTCGGCACCAACCTCTCG
AGCGACAGCGCATCAATGCAGTAA

SEQ ID NO: 20, ABJ97680, hypersensitive response-functioning factor A [Xanthomonas oryzae pv. oryzae]

MNSLNTQFGGSTSNLQVGPSQDTTFGSNQGGNQGI SEKQLDQLLCQLISALLQSSKNAEEGKGQGG
DNGGGQGGNSQQAGQQNGPSPFTQMLMHIVGEILQAQNGGGAGGGGGFGGGFGGDFSGDLGLGTNLS
SDSASMQ

SEQ ID NO: 21, gi|42717988:1136-1555 Xanthomonas oryzae pv. oryzae hrp gene cluster, partial sequence

ATGAATTCTTTGAACACACAATTCGGCGGCAGCACGTCCAACCTTCAGGTTGGCCCAAGCCAGGAC
ACAACGTTTCGGTTCGAACCAGGGCGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGTTGCTG
TGCCAGCTCATCTCGGCCCTGCTTCAGTCGAGCAAAAATGCTGAGGAGGGTAAGGGTCAGGGTGGC
GATAATGGCGGTGGCCAGGGCGGCAATTCGCAGCAGGCTGGGCAGCAGAATGGCCCCCTCGCCATTC
ACCCAGATGCTGATGCATATCGTCGGAGAGATTCTCCAGGCGCAGAATGGTGGTGGTGGTGGTGGC
GGCGGGTTCGGCGGCGGGTTCGGCGGTGACTTTAGTGGCGACCTCGGCCTCGGCACCAACCTCTCG
AGCGACAGCGCATCGATGCAGTAA

FIGURE 4 (continued)

10/96

SEQ ID NO: 22, AAC95121.2| Hpa1 [Xanthomonas oryzae pv. oryzae]

MNSLNTQFGGSTSNLQVGPSQDITTFGSNQGGNQGISEKQLDQLLCQLISALLQSSKNAEEGKGQGG
DNGGGQGGNSQQAGQQNGPSPFTQMLMHIVGEILQAQNGGGAGGGGFGGGFGGDFSGDLGLGTNLS
SDSASMQ

**SEQ ID NO: 23, gi|50428340:1138-1557 Xanthomonas oryzae pv. oryzae
hrp gene cluster, complete cds**

ATGAATTCTTTGAACACACAATTTCGGCGGCAGCACGTCCAACCTTCAGGTTGGCCCAAGCCAGGAC
ACAACGTTTCGGTTCGAACCAGGGCGGCAACCAGGGCATCTCGGAAAAGCAACTGGACCAGTTGCTG
TGCCAGCTCATCTCGGCCCTGCTTCAGTCGAGCAAAAATGCTGAGGAGGGTAAGGGTCAGGGTGGC
GATAATGGCGGTGGCCAGGGCGGCAATTTCGCAGCAGGCCGGGCAGCAGAATGGCCCCCTCGCCATTC
ACCCAGATGCTGATGCATATCGTCGGAGAGATTCTCCAGGCGCAGAATGGTGGTGGTGGTGGTGGC
GGCGGGTTCGGCGGCGGGTTCGGCGGTGACTTTAGTGGCGACCTCGGCCTCGGCACCAACCTCTCG
AGCGACAGCGCATCGATGCAGTAA

SEQ ID NO: 24, BAD29979, Hpa1 [Xanthomonas oryzae pv. oryzae]

MNSLNTQFGGSTSNLQVGPSQDITTFGSNQGGNQGISEKQLDQLLCQLISALLQSSKNAEEGKGQGG
DNGGGQGGNSQQAGQQNGPSPFTQMLMHIVGEILQAQNGGGAGGGGFGGGFGGDFSGDLGLGTNLS
SDSASMQ

**SEQ ID NO: 25, gi|82393799:1-378 Xanthomonas oryzae pv. oryzicola
hpaGXooc gene, complete cds**

ATGAATTCTTTGAACACACAATTTCGGCGGCAGCGCGTCCAACCTTCAGGTTGACCAAAGCCAGAAC
GCGCAATCCGATTCGAGCCAGGGCAGCAATGGCAGCCAGGGTATCTCGGAAAAGCAACTGGACCAG
TTGCTGTGCCAGCTCATCCAGGCCCTGCTTCAGCCGAACAAAAATGCTGAGGAAGGTAAGGGTCAG
CAGGGTGGCGAGAATAATCAGCAGGCCGGGAAGGAGAATGGCGCCTCGCCACTCACCCAGATGCTG
ATGAATATCGTCGGAGAGATTCTCCAGGCGCAGAATGCCGGCGGCAGCAGCGGCGGCGACTTTGGT
GGCAGTTTCGCCAGCAGCTTCTCGAACGACAGCGGATCGATGCAGTAA

**SEQ ID NO: 26, ABB72197, hpaGXooc [Xanthomonas oryzae pv.
oryzicola]**

MNSLNTQFGGASNFQVDQSQNAQSDSSQGSNGSQGISEKQLDQLLCQLIQALLQPNKNAEEGKGQ
QGGENNQQAGKENGASPLTQMLMNIVGEILQAQNAGGSSGGDFGGSFASSFSNDSGSMQ

**SEQ ID NO: 27, gi|21112286:70-435 Xanthomonas campestris pv.
campestris str. ATCC 33913, section 131 of 460 of the complete
genome**

TCAGGCTTGGCCGGTGATGCTCGACAGGTTGGCATTGAAGCCGCCACCCAAGCTGGTGCCGCCCAT
GCCGGCGCCGCCTTGGTTCTGCATCAGCTGCATCACGATCTGCATCAGCATCTGCGTCAACGGACT
CACACCGTCCTGTTGACCGCTCTGCGGTTGTTTGTCTCCGCACTCCTGATCGGCATCGCTGCCCTG
GCTCTGTTGGAGCATCATCATGATGAACATGGCGAGCAGCTGATCCAGCTGCTGCTCGGAGTCAGC
CGAAGGCGAGCGCTGACTGGAGTTCTGGGTTTGTCTGGGGCCCGATGCCCATCGTCTGCAGGTTGAT
GAAGTTGGAAAATTTGTTTCCGATAGATGAGTCCAT

**SEQ ID NO: 28, AAM40538, Hpa1 protein [Xanthomonas campestris pv.
campestris str. ATCC 33913]**

MDSSIGNKFSNFINLQTMGIGPQQTQNSSQRSADSEQQQLDQLLAMFIMMMLQQSQGSADADQECG
DEQPQSGQQDGVSPLTQMLMQIVMQLMQNQGGAGMGGTSLGGGFNANLSSITGQA

FIGURE 4 (continued)

[illegible]

FIGURE 6

12/96

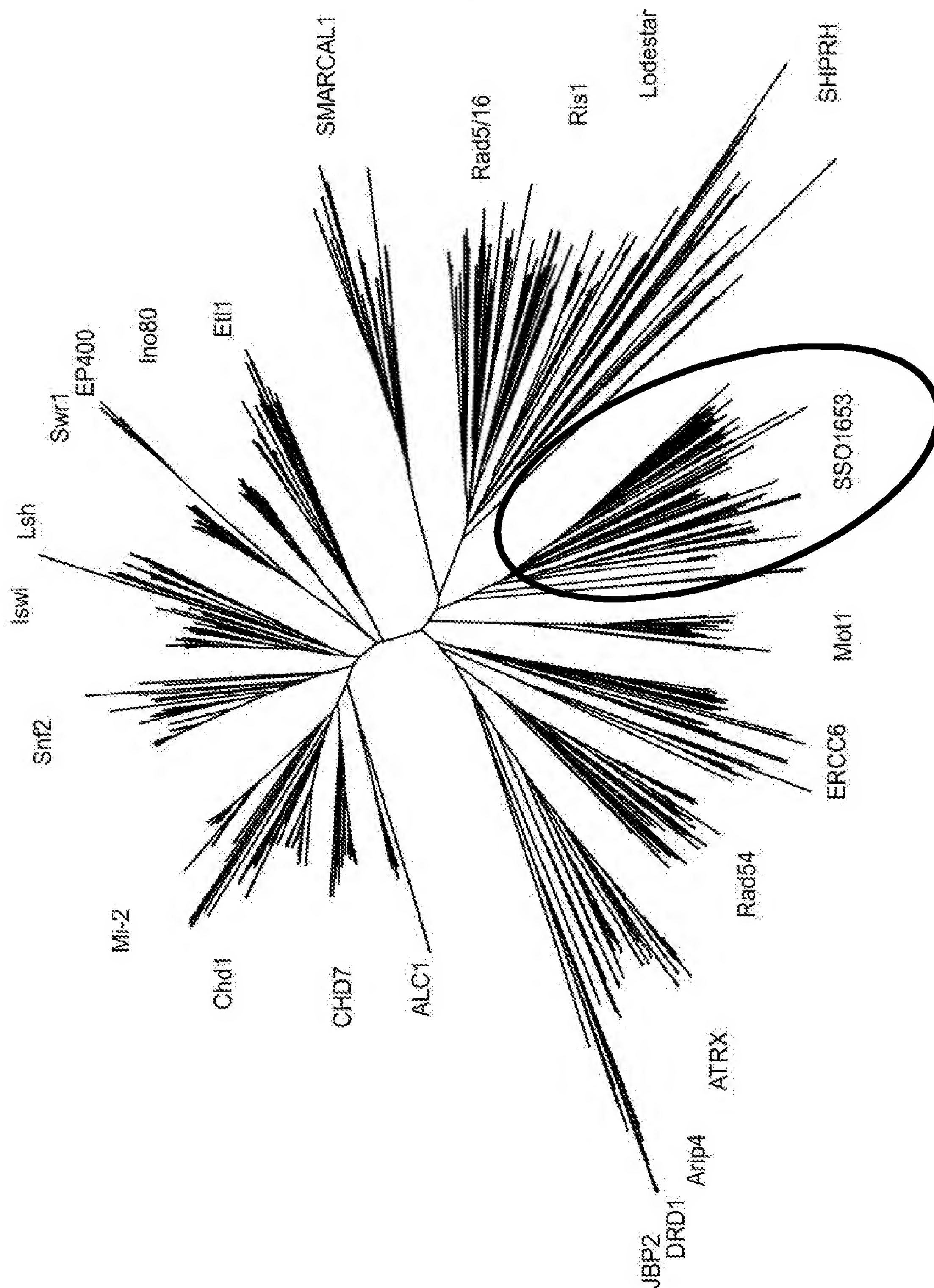


FIGURE 7

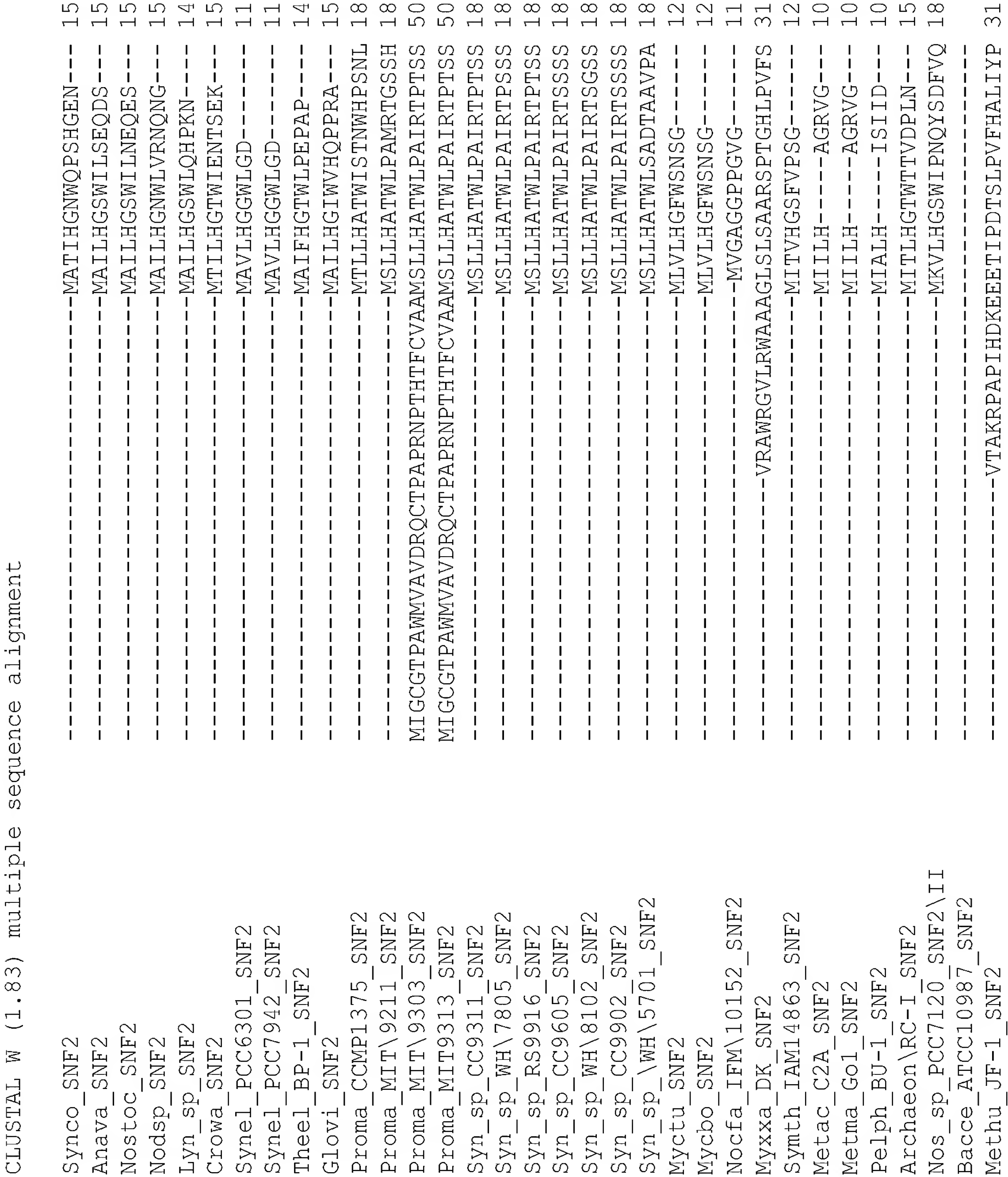


FIGURE 8

Synco_SNF2	-----GGKLFLWADTWGHPLPETIG-----DRHPFALDLPDLLQAWSN	53
Anava_SNF2	-----YLFIWGETWRSPQVNFSEF---EEIALNPLALSASELSEWLQS	54
Nostoc_SNF2	-----CLFIWGETWRSPQVDFNF---AEISLNPLALSASELSEWLQS	54
Nodsp_SNF2	-----CLFIWGETWRSSRVDFALNVSQDIPLHPLVMSPIDLSELLSY	57
Lyn_sp_SNF2	-----YLFIWGETWRR-ITPNEFNPADGVLGYPFALSPVELEKWCSE	55
Crowa_SNF2	-----HFFIWGETWRSLSDDISSD--DSILMYPFSVDKQGIIEQLNS	55
Synel_PCC6301_SNF2	-----RFCVWAEAWQAGEPQSAAEIAIH---PYAIAATDLNDWCQK	49
Synel_PCC7942_SNF2	-----RFCVWAEAWQAGEPQSAAEIAIH---PYAIAATDLNDWCQK	49
Theel_BP-1_SNF2	-----QFFIWAEEWRS-----LAQAIT-----PWAPPAIPVYPYATQ	46
Glovi_SNF2	-----G--LFLWGETWRQVAKRRKRS--EAPAPHYVQQPAELSPRLAA	55
Proma_CCOMP1375_SNF2	G-----QSELFLWADQWRVVTPKQIIQ---TPSPHPFSLSSDELKEWLNS	60
Proma_MIT\9211_SNF2	N-----PG-LLIWADSWRVAKPSIVSN---QPVIHPFALSADLRIWLLQ	59
Proma_MIT\9303_SNF2	G-----RPALLVWADTWRVATPAGPAA---TPALHPFTLNPDDLRAWLIE	92
Proma_MIT9313_SNF2	G-----RPALLVWADTWRVATPAGPAA---TPALHPFTLSPDDLRAWLIE	92
Syn_sp_CC9311_SNF2	G-----RAALLVWADTWRVAEAPAGPST---TPALHPFTLSPDDLRAWLIE	60
Syn_sp_WH\7805_SNF2	G-----RAALLVWADTWRVADPLGPGA---TPALHPFTLSAEDLRAWLTE	60
Syn_sp_RS9916_SNF2	G-----RAALLVWADTWRVAEAPAGPV---TPATHPFTLSADDLRAWLSE	60
Syn_sp_CC9605_SNF2	G-----QPALLVWADTWRVATPEGPGL---TPALHPFTLSHEDLRAWLSE	60
Syn_sp_WH\8102_SNF2	G-----QPALLIWADTWRVATPEGPGL---TPALHPFTLEPDDLRAWLQE	60
Syn_sp_CC9902_SNF2	G-----QPALLIWADTWRVASPEGPGL---TPALHPFTLGSDDLRAWLTE	60
Syn_sp_WH\5701_SNF2	LGGG-YRPGLLWADTWRVAEPOTPAS---EAPQHPLSLDQDDLGAWLEE	64
Myctu_SNF2	-----GMRLWAEDSD-LLVKSPSQA-----LRSARPHPFAAPAD	45
Mycbo_SNF2	-----GMRLWAEDSD-LLVKSPSQA-----LRSARPHPFAAPAD	45
Nocfa_IFM\10152_SNF2	-----ATCLDGRMLHGLWSPGGLV-----LWTEGEVPPALP--	43
Myxxa_DK_SNF2	GFSVATDVGVLFAGLSVRALVHQPGGG-----PLRAPHGQPGRPAA	73
Symth_IAM14863_SNF2	-----ASGFFFLWGLDGVAAARDAAPPG-----RRRRGVPRHPCA	46
Metac_C2A_SNF2	-----KQFFLWGESPAENETPPVRRGRKPKKPVAKPYPYDSGVENLSS	53
Metma_Go1_SNF2	-----KQFFLWGESPAENETPVVRRGRKPKTPIVKPYPYDSGFENLSS	53
Pelph_BU-1_SNF2	-----GVPLLWSEGKKIGMLKELRLAT-----AGIGMFS-	39
Archaeon_RC-I_SNF2	-----GTFFLWGESD--PATQHKRRGRPRKRSAGEKHPPFHAGIKELEA	56
Nos_sp_PCC7120_SNF2\II	S-----GAFYLWVETPINNKKRTHQVHPGHLSSLELLNFLTQTLGIKE	62
Bacce_ATCC10987_SNF2	-----MINQTEVTIRLQHVSHG-----	17
Methu_JF-1_SNF2	AVEG---VAICAHEYITDKPAPVRKKGYAKDKPGEYPYSLDHTALKTLIEN	78

FIGURE 8 (continued)

Synco_SNF2	LPLAFPKADGVT-----EAALTHLPSHRQK-	80
Anava_SNF2	QHQAIAQILPQQ-----LAKKTSKAASSPTNLPISQIIVLPTEISQPR	99
Nostoc_SNF2	QHQAIAKLLPQQ-----LEKRTSKAASSVKINLLTHSQIIALPTEISQPR	99
Nodsp_SNF2	HNIKIPSLIQSQVALSGTGRTRKSTSTTKFSWTTTHSLIIDLPTHISENN	107
Lyn_sp_SNF2	KQLSIESKVVT-----ETLALPTKLSP--	78
Crowa_SNF2	NKIKIEKNKNIES-----VSQIFYLPSKFIK-	82
Synel_PCC6301_SNF2	YRLGS-LTGTP-----EVLLSIPSDLKKE-	73
Synel_PCC7942_SNF2	YRLGS-LTGTP-----EVLLSIPSDLKKE-	73
Theel_BP-1_SNF2	RKTPLRKTARPS-----ATYVALPAQIQGH-	71
Glovi_SNF2	QFPQIPLSLLVP-----ETLALQLPATVEN--	80
Proma_CCMP1375_SNF2	KKLLPNESINTS-----ACLTLPSPKPIH--	83
Proma_MIT\9211_SNF2	KKLLPKESIECT-----ALLTLPKSIKNS	84
Proma_MIT\9303_SNF2	RDLLPDEIIDAT-----ACLTLPSRTVKPR	117
Proma_MIT9313_SNF2	RDLLPDEIIDAT-----ACLTLPSRTVKPR	117
Syn_sp_CC9311_SNF2	RDLLPDGIIDAT-----ACLTLPSSRSVKPR	85
Syn_sp_WH\7805_SNF2	RDLLPDGIIDAT-----ACLTLPSSRSVKPR	85
Syn_sp_RS9916_SNF2	RELLPDGIIDAT-----ACLTLPSRTVKPK	85
Syn_sp_CC9605_SNF2	RDLLPGGCIDAT-----ACLTLPSRTVKLR	85
Syn_sp_WH\8102_SNF2	RDLLPGGSIDAT-----ACLTLPSRTVKPR	85
Syn_sp_CC9902_SNF2	RDLMPGGSIDAT-----ACLTLPSSRSVKPR	85
Syn_sp_WH\5701_SNF2	ADLWTEDFRPAG-----ATLCLPSRRQGAR	89
Myctu_SNF2	LIAGIHGPKPAT-----AVLLLPSLRSAPL	70
Mycbo_SNF2	LIAGIHGPKPAT-----AVLLLPSLRSAPL	70
Nocfa_IFM\10152_SNF2	-----DPAG-----ALLRASRFRHR--	58
Myxxa_DK_SNF2	HGAGNPVQGRSQ-----ACLRVPLARTEFT	98
Symth_IAM14863_SNF2	TEP-----EALYPALRGLPYL	62
Metac_C2A_SNF2	ALELLLGSTG-----RKKAEEINVWIPTAG---	78
Metma_Go1_SNF2	ALELLLGSTD-----RKKAEGINVWTPTIG---	78
Pelph_BU-1_SNF2	----LLDNT-----TKEFCVWLPCRE---	56
Archaeon_RC-I_SNF2	GAGAINSSCIRHI-----ADAGARAEQVLILPSATDRPL	90
Nos_sp_PCC7120_SNF2\II	TEAQLKQRICSK-----YFALPTANNEP-	85
Bacce_ATCC10987_SNF2	-----WFLWGEDDDSGTP-	29
Methu_JF-1_SNF2	CFGAYDDLKATR-----WIIYLPAAEETVP-	102

FIGURE 8 (continued)

Synco_SNF2	-----IPLPFTGQDPVAMDAKYLHRSWQVTGVNLTPS	114
Anava_SNF2	-----IFISPVHSAALESADSE-VYLQWRVEGFCCLPPS	137
Nostoc_SNF2	-----ILISPVHSAALASESDSE-VYLQTRVEGFCCLPPS	137
Nodsp_SNF2	-----EFISPLHSATLGSEINSP-QYLQWRVEGFCCLNPT	145
Lyn_sp_SNF2	-----GLYPLQSTPQTDSETDSESICLYPWKIEGICLNST	115
Crowa_SNF2	-----QSIPLLSTELKDKDFEQGDIQLIAWKIEGIKLNVD	119
Synel_PCC6301_SNF2	-----AVLPFLSGQEIPDG-----ALLSWQIPVLSLEAA	103
Synel_PCC7942_SNF2	-----AVLPFLSGQEIPDG-----ALLSWQIPVLSLEAA	103
Theel_BP-1_SNF2	-----QLLPPP--LAEVQG-----ELLFLWQVPGWSIPAS	99
Glovi_SNF2	-----VVYSASIAPEG-----KLELEPWLVGEFGLDGH	109
Proma_CCMP1375_SNF2	-----KKNNQKSNQKTGIESEWKGLPLQAHEEIIATQYECWPWKVDGISLTTV	131
Proma_MIT\9211_SNF2	LDKKLNGVTDSONTSQDQWWSGLPLQAGEPVTKQCEWWPWQVEGIAIKPS	134
Proma_MIT\9303_SNF2	SKAKNVSTESDEDEDKHKTSWTGLPLQAGEPIPKQTEWWPWQVQGLAVEPA	167
Proma_MIT9313_SNF2	NKTKNVSTESDEAKDNKTSWTGLPLQAGEPIPKQTEWWPWQVQGLAVEPA	167
Syn_sp_CC9311_SNF2	KKR-----ETETSSTEQPSWTGLPLQAGEPIPKQTEWWPWQVQGLAIDPM	130
Syn_sp_WH\7805_SNF2	RPRG--SAAATPSSEEQPPWCGLPLQAGEPIPKTTEWWPWQVQGLAIEPM	133
Syn_sp_RS9916_SNF2	RKR-----GETAPVDEG-WTGLPLQAGEPIPKQTEWWPWQVQGLAVEPG	128
Syn_sp_CC9605_SNF2	KSRS-----TKEEPTPEPPGWTGLPMQAGEPIPKQTEWWPWQVQGLAVEPS	131
Syn_sp_WH\8102_SNF2	KSRS-----KTAEPAPPEEPIWTGLPMQAGEPIPKQTEWWPWQVQGLAVEPS	131
Syn_sp_CC9902_SNF2	KSRT-----QPSEPAPEGPAWTGLPMQAGEPIPKQMEWWPWQVQGLAVEPS	131
Syn_sp_WH\5701_SNF2	GKKK-----SDTSSWSGLPLQAGEPIPKSVVEWWPWWRVEGWLEPG	129
Myctu_SNF2	-----DSPELIRLAPRPAAR--TDPMLLAWTVPVVDLDPT	103
Mycbo_SNF2	-----DSPELIRLAPRPAAR--TDPMLLAWTVPVVDLDPT	103
Nocfa_IFM\10152_SNF2	-----AQVLVPGPAG-----PQLT--QVRAHALVPQ	82
Myxxa_DK_SNF2	-----FAAMPLVFLPDAETLFLWGPDRILPRELAGLPETGD	133
Symth_IAM14863_SNF2	-----NTLSLVQWQPGPD--GVSPARVPGIALSVPN---	91
Metac_C2A_SNF2	-----WNPISSPLVAEIPASKAEISLAPWTVHAYPLEAE	113
Metma_Go1_SNF2	-----GNPVPSSPLVAEISDSKAEPALAPCTVHAYPLEAE	113
Pelph_BU-1_SNF2	-----KKAVPSSPLVGAMPDLSDEEQQLHAFPTALRLNFN	91
Archaeon_RC-I_SNF2	-----RSASPSALESGEETNPDSLSLQFLPWTVTGINIKPG	125
Nos_sp_PCC7120_SNF2\II	-----LPSPELVKYLEVEVPEEYENFQYWQVTCYETVTS	119
Bacce_ATCC10987_SNF2	-----LSVTSWKRN-----AFTWHSTSYFG	49
Methu_JF-1_SNF2	-----PSSQFSSKKKPSPEKKKLPLVPMYIPVLLCPYE	135

FIGURE 8 (continued)

Synco_SNF2	QTLTLQSIPLGG---QALAN-----LGSEFFYFGQL	143
Anava_SNF2	AAVKFLTSLPLNI-----TSTENAF-----LGGDLRFWSQI	168
Nostoc_SNF2	AAIKLLTSLPLNI-----TSGENAF-----LGGDLRFWSQI	168
Nodsp_SNF2	EAIKFLAAVPLNA---AREEDTL-----FGGDLRFWSQI	176
Lyn_sp_SNF2	EAFDFLQSLPLGN---LTSENSF-----IGSDLQFWSHL	146
Crowa_SNF2	DTNILSQPLPLGL---TNNDENY-----IGDNLKFWTHI	150
Synel_PCC6301_SNF2	IAGQWLATLPLG-----SAEDHPW-----LGPDLRFWSHI	133
Synel_PCC7942_SNF2	IAGQWLATLPLG-----SAEDHPW-----LGPDLRFWSHI	133
Theel_BP-1_SNF2	EVLEQLHQLSL-----HGQDSGS-----IGDDLRYWLHV	128
Glovi_SNF2	QAFELLGVPPLG-----GGDAS-----IGDDLRFWSQC	137
Proma_CCMP1375_SNF2	EATEWLTKLPLS-----KKDSD-----LSEELLWWAHL	159
Proma_MIT\9211_SNF2	EAASWLANLPLT-----KKDPE-----LSEELIWWSHL	162
Proma_MIT\9303_SNF2	AATAWLSKLPLS-----GDHPD-----LADELRWWSHL	195
Proma_MIT9313_SNF2	AATAWLSKLPLS-----GNHPD-----LADELRWWSHL	195
Syn_sp_CC9311_SNF2	AATAWLSKLPLS-----GRHPD-----LADELRWWSHM	158
Syn_sp_WH\7805_SNF2	AATAWLAKLPLS-----GHHPD-----LADELRWWSHM	161
Syn_sp_RS9916_SNF2	AATAWLARLPLS-----GRHPD-----LADELRWWSHM	156
Syn_sp_CC9605_SNF2	AATEWLSRLPLS-----GTNPD-----LADELRWWSHL	159
Syn_sp_WH\8102_SNF2	AATEWLSRLPLS-----GRNPD-----LADELRWWSHL	159
Syn_sp_CC9902_SNF2	AATEWLARLPLS-----GRHPD-----LGDELRWWSHL	159
Syn_sp_WH\5701_SNF2	AATLWLGRPLS-----GDHPD-----LADDLRWWSHL	157
Myctu_SNF2	AALAAFQDP-----APDVR-----YGASVDYLAEL	128
Mycbo_SNF2	AALAAFQDP-----APDVR-----YGASVDYLAEL	128
Nocfa_IFM\10152_SNF2	AAVDVLRQR-----LPVES-----VAGDLRFLAHV	107
Myxxa_DK_SNF2	RASALLVTPEGLRECEGHGLPLAATVERLAVVQTSEAESFPGSIALWTLA	183
Symth_IAM14863_SNF2	AVQWLLDLPDHR-----GTPLRP-----GHSIQLWCVA	120
Metac_C2A_SNF2	EAIVLLCACMGKK-----VLAPG-----IISGNDLLWWADA	144
Metma_Go1_SNF2	EAIVLLCTCMEKK-----VLAPG-----IISGNDLLWWADA	144
Pelph_BU-1_SNF2	ALFELSLLTEKGN-----IPCSG-----IIFGSSLHWARQV	122
Archaeon_RC-I_SNF2	NALVLLSSIAESQ-----KRIGD-----MAIGPDLLYWSKV	156
Nos_sp_PCC7120_SNF2\II	VKAVIAINIIKLLKDIHFLALYNAS-----EFQLGSDLLFWYHY	158
Bacce_ATCC10987_SNF2	TFLKEASFEGRQG-----VMLTNAQAFFEYI	74
Methu_JF-1_SNF2	TFFQIWKAAQNTD-----KNYIAGDSFQYI	160

FIGURE 8 (continued)

HRWCLDLVLRGKFVPGLEQRGED-GNYAQAQWIPILDSIQDQTHLAQFSQR 192
ARWSLDLISRSKFLPIIQRPN--NSVSAKWQVLLDSAVDGTREKFAAK 216
ARWSLDLISRSKFLPIIQRPN--NSVSAKWQVLLDSAVDGTREKFAAK 216
ARWSLDLISRSKFLPTIQRFD--SSIVARWQVLLDSAIDGTREKFSK 224
SRWSLDLLARSKFLPSLTFNPSK-DHFIAEWKPLLLDSATDQARLIRFSKQ 195
YRWSLDLLTRGKYLPMQEEQDN---NCYGQWEPLLLDSLVDQQRFSKFIQT 197
YRWAQSLIARGRFYPALESSDR---GLTAVWLPLFNQAGDRQRFDRYSQQ 180
YRWAQSLIARGRFYPALESSDR---GLTAVWLPLFNQAGDRQRFDRYSQQ 180
SRWLLDLIVRGQYLPTEG-----WRILLTHGDRDRRLRHFSQL 167
ARWVLDLLVRAKYLPDLESGDGQ-EIPTARWVPLLLDSAVDQARLLKEFAAR 186
ERWSLNLIASGLWLPOVKLHKKEGNEYRASWIPLLNQENERNRLEEFKN 209
ERWSLSLIARGLWLPOVELNTIDNIGARARWSPLLNNENENKRLFEFSIR 212
QRAWLSMIARGRWLPQVELSKGEGYPHRARWTPLLNREDDRRRLEDLAAQ 245
QRAWLSMIARGRWLPQVELSKGEGYPHRARWTPLLNREDDRRRLEDLAAQ 245
QRAWLSLVARSRWLPOVELSKGEGYPHRARWVPLLNREEDRRRLEDLAAQ 208
QRAWLSLVARGRWLPOVELSRGEGYPHRARWVPLLNREEDRRRLEDLAAAR 211
QRAWLSLIARSRWIPQVELSKGEGYPHRARWVPLLNREDDRRRLEDMAAR 206
QRAWLSLVARGRWIPQMEFSKGEGYPHRARWVPLLNREEDRRRLEDLAAAS 209
QRAWLSLVARGRWIPQMEFSKGEGYPHRARWVPLLNREEDRRRLEDLAAAS 209
QRAWLSLVARGRWIPQMEFSKGEGYPHRARWVPLLNREEDRRRLEDLAAAT 209
QRAWLSLVARGRLLPQVEGG-----RARWLPLINREDDRRRLEDLASR 200
AVFARELVERGRVLPQLRRDTH---GAAACWRPVIQG-RDVVAMTSLVSA 174
AVFARELVERGRVLPQLRRDTH---GAAACWRPVIQG-RDVVAMTSLVSA 174
ADGIDRWVRAGRVPDLHRADG---QWWARWRLVGGARQRAWLAELAVA 153
SKLLELVARERVVPTLLRRGE---RIEARWAAALSAATEDAGRVAALARS 230
SKLLEFLGRGLMPVIAEAG---VLSAGWALHLLTDADDVRRITRLAAG 167
LKFAGSLVAGQKYLPGVRGEG-----EYKAFWEVPVFSGEDAGELARLAKQ 190
LKFAGSLVAGQKYLPGVRGEG-----EYRAFWEVPVFSGEDAGKLAKLAKQ 190
VKIALNIVRTQSLPSIIKNDT-----FWEALWLPLPDSATSLAVEQLADA 168
AKFTLKLIIISQQFRPEVVEVMMSG--KAYSRRWFALTDETDKHYASLENS 204
TQSFQRQIITKDQYIPSLKYRAN-AATTKKKPKQPPPGFEIYAGWEIISEQ 207
ANKPMNSFARIQMNGPITALTEDANELWDAFTSGSFVPDMERWPKQPSWK 124
S-ILMESTVRLLIQNGRFKPSLERTTFAGYHAVWVPALSPQDMEWVSDFSSR 209

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon_RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

FIGURE 8 (continued)

VPACALANLT-----DSQEPQMLVVD 213
MPLVCRITYQR-LGNEEL-----SPSP-----IYIDFPSQPQELILG 251
MPLVCRITYQE-IGS-----GESP-----IYIDFPSQPQDLILG 248
MPLACRTYRKMGSGEWGVSGEESSPSI-----MYVDFTFEPQELLLG 268
IPSACRIYQLWSKEAQN-----QFEN-----LALDLPQNPNLIDD 231
MPNSSLAYHNLMEG-----ELSSLLKQTTILD 225
LPFSQFCYQAIETA-----AACPWQPQPQDILLR 209
LPFSQFCYQAIETA-----AACPWQPQPQDILLR 209
MPDLCRCYQADGTA-----LQLP--PHAADLLAD 194
LPGACRAATP-----ELSPHQILKS 206
IPLVAICAVPWIEAKG---QIVNTEQVSNNNNTLSLYRPRHNRVEVMD 255
LPLVATCAIKREETSEENQNHILKTTPRETLDEYGLAVCRPINSRLQVAY 262
LPLVATCALPWREPTGRRSNRMTRLRPEAMRAANPVASCRPRSGLRVAS 295
LPLVATCALPWREPTGRRSNRMTRLRPEAMRAANPVASCRPRSGLRVAS 295
LPLVATCALPWREPTGKRSNRITRLRPEAMRAANPVACCRPRSGLRVAT 258
LPLVATCALPWREPTGKRSNRITRLRPEAMRAANPVACCRPRSGLRVAT 261
LPLVATCALPWREPTGKRSNRITRLRPEAMRAANPVACCRPRSGLRVAT 256
LPLVATCALPWREPLGRRSNRTTTLRPEAMRAANPVASCRPRSGLRVAT 259
LPLVATCALPWREPMGRRSNRMTRLRPEAMRAANPVACCRPRSGLRVAT 259
LPLVATCALPWREPLGRRSNRTTTLRPEAMRAANPVACCRPRSGLRVAT 259
LPQVAVAAL-----EPGQGEAGVAMACWRPFGSGRRRLAS 234
MPPVCRAEVGG-----HDPHELATSALDA 198
MPPVCRAEVGG-----HDPHELATSALDA 198
MPAALRVAG-----QPAAVLDDDLVTE 174
MPPGAHAVPAGARPG-----RAVWAPDALLRA 257
LPEACRALVPPDRTPN-----TYPLPVADGLVHQ 196
MPPAAKALALETSSVQP-----EILAAVAARQ 217
MPPAARALAPEASSMPP-----EMPAALAAKQ 217
MPAVCRSLG-RTDTQPP-----ETPKKLLKKG 194
MPLACIAVSGKAGIYN-----RKEALDL 227
YEANIQKYIEYMPPLICVAG-----NSTQTDKLEFFAPET 241
VQNTP-----IEDETLAS 137
MPTVCKYAIIPRVAKDPY-----IYKPETRLEK 236

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Mycctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon_RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

FIGURE 8 (continued)

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Mycu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon\RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

255
300
297
317
280
266
250
250
235
252
297
304
337
337
300
303
298
301
301
276
244
244
220
301
243
266
266
239
269
285
181
280

LIQKLLQAQIGAVSPS-----LANVKEVWLNWLRGLTHGGQTS LGT
FLNSAIDTQLREMVGNQPVV-ETRLMASLPSAVRQWLQGLSGASNSVDAD
FLNSAIDTQLREMVGNQPVV-ETRLMASLPSAVRQWLQALIAASNSIDAD
FLNSTIDAQVREMLASQPLL-ETRVMASLPSAVRQWLQGLTSASHTVNAD
FLTAIIDSQVKKVAEESEKK-AITNLTAIQPIVQSWLHALASESNLAKSK
FLSTIINQVRQFID-----VAITPSSFIQKWLYSLTQDLSKFEAS
VLQTWLTARLQPAIAA-----GTLVSADLLAAWQQSLAN-GKPLKLE
VLQTWLTARLQPAIAA-----GTLVSADLLAAWQQSLAN-GKPLKLE
FLQHTLQGYLHTALAD-----LELPKVGLAKEHGHWLA F-LKTGQTP
FLSAMLDAVRVRTLACEP----PDPRTLPA GAVRPWLLALAH AQPLKSP
LLEE LIDAQLRKDFQP-----RTKNLDPLLKAWQEALG TKDGIINLS
LLEE LVDGQLRKDFEE-----SSEDLDPLLKAWQEALGSHNGVIRLP
LLEE LLD AQLRTGFEA-----SEQGLDPLLTA WQEALGSDSGVINLP
LLEE LLD AQLRTGFEA-----SEQGLDPLLTA WQEALGSDSGVINLP
LLADLMDAQLRKGF TP-----DPDGLDPLLRA WEEALSSDTGEIQLS
LLEDLVDAQLRKGFHP-----DDEGLDPLLCA WENALSS ETGVIDLN
LLEDLVDAQLRTGFTA-----QTDGLDPLLAA WEEALGSDTGVIHLG
LLEDLVDAQLRKDFEP-----STDGLDPLLTLWQEALGSETGVIEIG
LLEDLVDAQLRKDFEP-----STDGLDPLLTLWQDALGSETGVIEIG
LLEDLVDAELRKGFEP-----TTEGLDPLLTLWQEALASETG VVEVG
ILTHLVDA RM RAGFTP-----SEEGLDPLLAA WQALGPGDGRLLDLG
MVDAAVRAALSPMDLLPPR---RGRS-KRHR AVEAWLTALTCPDGRFDAE
MVDAAVRAALSPMDLLPPR---RGRS-KRHR AVEAWLTALTCPDGRFDAE
LTDPIVTR LADAPVTHPL---VRAL-VRDQPLETGSHQLAEVLRWRRES
FLDATVD AFVRAARGAPSL---PARR--AASWDERWREALTGAR-RDFAP
FMRTAAAGVIRLLLEEEPL---PEAQSLQDTALRHWLAALTGA EARDLPP
FIEEALDWI VRSEIGE KE LAKEARKRKS FDSVHDAWVSALKSPD-GLIHG
FIEDSLDWI VRSEIGE KKLAKETRKRKSFDSVHDAWVSALRSPE-GLIYG
LLSFLVNTLSRTFERAG-----VPKISDFESI HD A WLHALSNSDPRLKWK
FINTALDTFIRDQIALP-----ADSRMTNLLSQAWLDSLG TGE--SIRL
LLRHFSEYLLNNLVSKTP-----LTA AF EKQIDDSL IHYCLYPQKHNP L
LFSAAVNESILQDNRSND-----GWEDAKRLYEHYDFTKRQLDAALHEE
FIVEMMRV IIR TALGGYT-----LKEETDPFFYEPSENE M QFMTDLLGVT
.

FIGURE 8 (continued)

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon_RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

S--KALQRLATSLDHWYLPVQNYLGQKNNQALAQRWGALRLQPPADDG
A--VGLERLEAALKAWTMPLQYQLASK-----NQFRTCFELRSPEPG-
A--VGLERLEAALKAWTMPLQYQLASK-----NQFRTCFELRSPEPD-
A--MEVERLEAALKSWTMPLQYQLVGK-----PSFRACFQLLPPASG-
K--SESKTLEKILSNWTAPLQQTAEH-----NLFRFTGFRLSPENN-
E--VERKGLKNAINNWKSSLSEYIIKSDNQPLGINQFRCFKLENPAKSG
D--SEASRLQTAIDRWLLP--VQNGAA-----QAWRMVLRVLPPTAQ-
D--SEASRLQTAIDRWLLP--VQNGAA-----QAWRMVLRVLPPTAQ-
E--LPPP-LIERLHRWQEPYREQHLR-----PQWRLALQLVPPDTA-
D--PETPALAEALATWRAPLSYQVRSR-----TCFRLQPPES-
N--ENAKRLEKASKNWKRGGLSSNVQPA-----KTCLELIAPIDD-
L--EDCERLAKASKNWKENLSGNVKA-----RACLELFAPLEG-
D--EEAERLATAASNHWREGVAGNVAPA-----RACLELFTPGEG-
D--EEAERLATAASNHWREGVAGNVAPA-----RACLELFTPGEG-
D--EETERLATAASNHWREGVAGNVAAA-----RACLELATPAD-
D--EDAERLATAASHHWREGVAGNVAAA-----RACLELATPNEG-
D--EDAERLATAASHHWREGVAGTVAAA-----RACLELETPDDG-
D--EEAERLATAASHHWREGIAGDFAAA-----RTCLELHTPPDG-
D--EQAERLASASFHWREGIAGDFAAA-----RTCLELQTPAEG-
N--EDAERLTAASLHWREGIAGGFAAA-----RTCLELNTPNEG-
D--DDCERLQVATHHWREAVAGRVEPA-----RACLELDTPDEG-
P--DELDALAEALRPWDDVGIGTVGPAR-----ATFRLSEVETENEETPA
P--DELDALAEALRPWDDVGIGTVGPAR-----ATFRLSEVETENEETPA
LTVDEPELVLRLLLEPDGETGIDGDG-----GDDRDDDTVA
EGFAERSVVDELTR-WSEPALGAR-----DKLRACFRLEPPTEER
GLPGAQELYAALDR-WSAPATGVLS-----HASLRTGVRLHLPGET
EE-KELLQLAFRTREWQRPLTVLTSP-----FRFCFRLEEPAAEE
DE-NELLQLAARTREWQRPLTILTSP-----FRFCFRLEEPAAEE
NE-QEIEQFACQLNAWRRPIDLHERSP-----FRFCIQLTEP----
SA-PEMKKDKDSAGRWTSMKTESKQA-----LKTCFILEPPAP--
KTHALTAEYQQWLGWKNRIIRTQAESP-----FHLCFQLHSPDAEQ
DWLRKIGYIEDDL-----PFTIGLRLQEPQE--
DPIRNKGFERTFLRAMQDWLTFSSSGRF-----APFEFCMIKDPPEG-

FIGURE 8 (continued)

Synco_SNF2	-----GGTWQLDYGLQALDDG	319
Anava_SNF2	-----ETEWTLAYFLQAADNP	356
Nostoc_SNF2	-----ETEWTLAYFLQAADDP	353
Nodsp_SNF2	-----ATDWILAYFLQAADDE	373
Lyn_sp_SNF2	-----QKNWTLDYCLQAIDEP	336
Crowa_SNF2	-----LEQSNWQLHYLQALDDP	334
Synel_PCC6301_SNF2	-----EQPWQLEFGLQAATDP	304
Synel_PCC7942_SNF2	-----EQPWQLEFGLQAATDP	304
Theel_BP-1_SNF2	-----DGDWHLAFGLQTEGET	290
Glovi_SNF2	-----QGEWKLHFLLLQTGDDP	304
Proma_CCMP1375_SNF2	-----LDLWDLNFSLQSESDP	350
Proma_MIT\9211_SNF2	-----EDLWDLQFSLQAEADP	357
Proma_MIT\9303_SNF2	-----EDLWELRFALQAEADP	390
Proma_MIT9313_SNF2	-----EDLWELRFSLQAEADP	390
Syn_sp_CC9311_SNF2	-----EDLWPLRFFLQAEADP	353
Syn_sp_WH\7805_SNF2	-----EELWDLRFYLQAEADP	356
Syn_sp_RS9916_SNF2	-----DDLWTLRFALQAEADP	351
Syn_sp_CC9605_SNF2	-----EDLWELRFGLQAEADP	354
Syn_sp_WH\8102_SNF2	-----EELWELRFGLQAEADP	354
Syn_sp_CC9902_SNF2	-----EELWDLKFGGLQAEADP	354
Syn_sp_WH\5701_SNF2	-----EDLWPLRFSLQAEADP	329
Myctu_SNF2	G-----SLWRLEFLIQSTQDP	303
Mycbo_SNF2	G-----SLWRLEFLIQSTQDP	303
Nocfa_IFM\10152_SNF2	-----LWRLEVCLRTEGEA	268
Myxxa_DK_SNF2	E-----PFVLSFHLQSPDDP	355
Symth_IAM14863_SNF2	D-----GEWELELTLHAPDEG	300
Metac_C2A_SNF2	ELEETEE-----SEAGKMDTKKGRKGIADIEVPEELWYVRYMLQSYEDP	350
Metma_Go1_SNF2	EIEETEETEEIEENEAGKRDTKKREGIADIEVPEGLWYVRYMLQSYEDP	356
Pelph_BU-1_SNF2	-----PLKGRK-----KERWHVAYQLQLKADP	297
Archaeon_RC-I_SNF2	-----DTEYPEAPWNLRyclQASDDP	328
Nos_sp_PCC7120_SNF2\II	-----IDNWQMQLVSSKKDP	342
Bacce_ATCC10987_SNF2	-----EFEMWKLETIVTPKRG	224
Methu_JF-1_SNF2	-----QTEPWDFTLAVRSEAEF	340
	: . :	

FIGURE 8 (continued)

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon\RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

EFLWPAASLWAMAGDRLVWQGRRV-DQGAESLLRGLGVAAQIYEPIAASL 368
EFLVDAGTIWQHPVEQLIYQORSI-QEPQETFLRGLGLASRLYPVIAPTL 405
EFLVDAATIWNPNVEQLIYQORTI-EEPQETFLRGLGLASRLYPVIAPTL 402
NLLVDAATIWHHPVEQLVYQNRTI-DQPQETLLRGLGLASRLYPVLTPSL 422
EFLVDAQTIWTHPVEAFVHNGRMI-KRPQETLLKGLGLASKLYPLLEPSL 385
NFLISAKVIWENPVTRLICNNRTI-NHPQETLLKGLGLASRLYYLIEESL 383
DRFRPASLLWQDPLPPGLPD-----QSQELLLRGLGQACRLYPQLQTSL 348
DRFWPASLLWQDPLPPGLPD-----QSQELLLRGLGQACRLYPQLQTSL 348
DTMLRAAEIWQCTQEALLYQGQVL-WQPQETLLRGLGLASRIYRPLDRSL 339
DSLMAAQQVWSSAG-----ELQEVFLAGLGLASRIFVPVERGL 342
SIRLAADQIWEAGVEVTKVGGITI-DNPSEILLLEGLGRSLEIFPPIEKGL 399
SLKVAEEAVWNADSAVLQIGDIQI-AQPGEILLLEGLGRALNIFQPIERGL 406
TIKVPAAAAWAAAGPKVLQIGEIRV-EHPGEVLLLEGMGRALTVFAPIERGL 439
TIKVPAAAAWAAAGPKVLQIGEIRV-EHPGEVLLLEGMGRALTVFAPIERGL 439
TLKLPAGAAWAAAGPSGQLGEIKV-EHPSEVLLLEGMGRALTVFQPIERGL 402
TLKVPAGAAWAAAGPEGLQGEIPV-EHPGEVLLLEGMGRALTVFQPIERGL 405
TLKVPAAALAWAAAGPKQLQGEIAV-EHPGELLLEGMGRALTVFPPPIERGL 400
SLKLPAAAAWAAAGAEPLQIGEIRV-DQPGEVLLLEGMGRALSVPFAIERGL 403
SLKLPAAAAWASGADQLQLGEVTV-EQPGEVLLLEGLGRALTVPFPIERGL 403
SLKLPAAAAWASGAETLQLGEIKV-DQAGEVLLLEGLGRALTVPFPIERGL 403
SLLLPAAGVWAAAGAGCLQGETEL-QQPGELLLEGLGRALQVFPEPIERGL 378
SLLVPAEQAWNDDGS---LRRWL-DRPQELLLTELGRASRIFPELVPAL 348
SLLVPAEQAWNDDGS---LRRWL-DRPQELLLTELGRASRIFPELVPAL 348
PAPVPATADPN-----LLR-----IAVEQLGRAQRAYPRLRDL 302
SLLVPAADVWKTRGRSLEKLGRAF-RDPQESLLEALGRAARLFPPALVL 404
ALPVTADAVWASLGAEVEIGGQRY-QGAEQRLLDLADLPAMARLFPPAPLI 349
SLLIPVKEAWKPK-KGSPLKRYDV-KNIRQFLLSSLGQAAGISAGIASSL 398
SLLIPVKEAWKPK-KGSPLKKYDV-KNIRQFLLSSLGQAASSISAGIASSL 404
SLLDAGDLWNPESEASQHALTYT-SDCTEFLLTSLGQASGLCPAVTQSL 346
SLVIPAETVWKELKKTLYLNKRY-DNPQEQLLQDLGKAMQMFPEIEPSL 377
SLKLALADYIWMNSKTKAGVHKEFGKDFDTNLLNLGYAARMYPKLWQGL 392
HRIYVYESIDSLPKRWH-----DYEERILET-----QESFSKLVFWL 261
SLLIPAEI IWELPDHQSGLFFPQA--AYLKHILLAGIGLLTSSSSSALWRPL 388
:

FIGURE 8 (continued)

Synco_SNF2	TERCPTGCGLD	AIQAYEF	ILAI	AHQ	LRDRGL	GVIL	PPGLERG	-GTAKRLG	417																									
Anava_SNF2	DTE	SPQFCHLN	PMQAYEF	IKAVAWR	FEDSGL	GVIL	PPSLAN	REGWANRLG	455																									
Nostoc_SNF2	DTE	SPQFCHLK	PMQAYEF	IKAVAWR	FEDSGL	GVIL	PPSLAN	REGWANRLG	452																									
Nodsp_SNF2	ETE	YPQCCRL	NPLQAYEF	IKSVAWR	FEDSGL	GVIL	PPSLTN	REGWANRLG	472																									
Lyn_sp_SNF2	QEAR	PQTCL	LTPLQAYEF	IKSNWR	FTDSGL	GVIL	PPSLVSQ	NGWANRLG	435																									
Crowa_SNF2	QDN	KPSFSEL	DP	IQVYEF	LR	SIANIL	KDNG	LVILPASLEQG	-VEEKRLG	432																								
SSynel_PCC6301_SNF2	ATAC	PEFHPL	TTAE	VYQ	LLKQ	VI	PQWQEQ	GIEVQLPPGLR	-GQGRHR	-LG	396																							
SSynel_PCC7942_SNF2	ATAC	PEFHPL	TTAE	VYQ	LLKQ	VI	PQWQEQ	GIEVQLPPGLR	-GQGRHR	-LG	396																							
Theel_BP-1_SNF2	QERS	PVAL	TLHT	TE	VYA	FLQSA	IAPLE	QQGVA	ILPPSL	RRNSAQHR	-LG	388																						
Glovi_SNF2	LVPQ	PTCCT	MST	VEAF	QFLK	AATWR	LRDSG	FGVLL	PE	SLADAGSL	RNLRLG	392																						
Proma_CCMP1375_SNF2	ESPT	PTH	TMKLS	ASEAF	VLIR	TAAAK	LRDMG	IVIL	PN	SLSG	--FASRLG	447																						
Proma_MIT\9211_SNF2	ENAT	PNNMQ	LT	PAEAF	VLRT	ASKQ	LRD	IGIG	VL	PR	SLSG	--LASRLG	454																					
Proma_MIT\9303_SNF2	DSAT	PEAMQ	LT	PAEAF	VLRT	AAAQ	LRD	VGV	VEL	PAS	LSGG	--LASRLG	487																					
Proma_MIT9313_SNF2	DSAT	PEAMQ	LT	PAEAF	VLRT	AATQ	LRD	VGV	VEL	PAS	LSGG	--LASRLG	487																					
Syn_sp_CC9311_SNF2	DSAT	PESM	Q	LT	PAEAF	VLRT	AVRQ	LRD	VGV	VDL	PP	SLSG	--LASRLG	450																				
Syn_sp_WH\7805_SNF2	DSAT	PEAMQ	LT	PAEAF	VLRT	AARQ	LRD	VGV	VDL	PP	SLSG	--LASRLG	453																					
SSyn_sp_RS9916_SNF2	DSAT	PEGM	Q	LT	PAEAF	VLRT	AARE	LRD	VGV	VEL	PAS	LSGG	--LASRLG	448																				
SSyn_sp_CC9605_SNF2	ESAT	PETM	Q	LT	PAEAF	VLRT	AARQ	LRD	AGV	G	VEL	PP	SLSG	--LASRLG	451																			
Syn_sp_WH\8102_SNF2	ETAT	PDTM	Q	LT	PAEAF	VLRT	AARQ	LRD	AGV	G	VDL	PP	SLSG	--LASRLG	451																			
SSyn_sp_CC9902_SNF2	ESAT	PETM	Q	LT	PAEAF	VLRT	AHQ	LRNAG	IG	VEL	PP	SLSG	--LASRLG	451																				
SSyn_sp_WH\5701_SNF2	DTAT	PERMAL	TP	PAEAF	VLRT	AALK	LRD	VGV	VV	LP	PS	LSGG	--LASRLG	426																				
Mycu_SNF2	RTAC	PSGLE	LDAD	GAYR	FLSG	TAAV	LD	EAG	F	GVLL	PS	WW	---DRRR	-KLG	394																			
Mycbo_SNF2	RTAC	PSGLE	LDAD	GAYR	FLSG	TAAV	LD	EAG	F	GVLL	PS	WW	---DRRR	-KLG	394																			
Nocfa_IFM\10152_SNF2	GDPH	SLDLL	LLPTE	VVAD	LV	AHGAQ	ALREAG	VR	LL	PP	RAW	---TIAE	PTLR	349																				
Myyxa_DK_SNF2	ESPR	PQALL	LEPD	TAWT	FLSE	GAR	VLSDAG	F	GV	IV	PGEL	TTSG	RRRLRLR	454																				
Synth_IAM14863_SNF2	RDP	APSR	MR	I	PADD	V	LALIQE	GAM	LLQ	QAGH	PVLL	PAALAKP	--AALRVG	397																				
Metac_C2A_SNF2	EAPN	PSGYS	LD	TK	EAYR	FL	TESA	AD	LSQAG	F	G	LL	PGWW	TRK	-GTKTHLK	447																		
Metma_Go1_SNF2	EAPN	PSGYS	LD	TK	EAYR	FL	TESA	AN	LSQAG	F	G	LL	PGWW	TRK	-GTKTHLK	453																		
Pelph_BU-1_SNF2	KKKQ	PGGF	DL	DT	E	GAYR	FL	LE	YAE	LL	RSAG	FV	VKL	PSWW	IGR	-RGNRIG	395																	
Archaeon\RC-I_SNF2	NTSK	PL	SAT	LS	TE	SEAYK	FL	TEA	AP	LLQ	DSGYS	I	IL	PEW	WRNS	-TGR	LKLG	426																
Nos_sp_PCC7120_SNF2\II	ETD	SPTGM	Q	LS	DEA	F	DLK	S	AW	V	LED	S	G	F	K	V	IPAW	YTPAG	RRRAKIR	442														
Bacce_ATCC10987_SNF2	KDGD	TFR	SEL	F	TEA	WN	FL	TEA	S	N	E	L	L	AAG	I	T	IL	LP	SWQ	NL	KATKPKLR	311												
Methu_JF-1_SNF2	SGSK	PTGGS	MT	LKEA	AT	FLG	S	D	L	A	R	R	K	G	V	T	V	L	P	D	W	W	T	D	T	T	Y	T	P	R	V	E	I	438

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon\RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

VKVVGEVQRQ-----RGQR-LTLQSLINYDLQIMMGSDNARLLTAKDFFA 462
LKISAETPKK-----KPGR-LGLQSLINFQWHLAIG-----GQTISKGEFDR 496
LKISAETPKK-----KPGR-LGLQSLINFQWHLAIG-----GQTISKAFFDR 493
LKISAETQKK-----KQGR-LGLQSLINFQWQLAIG-----GQTISKTEFNK 513
LSVQAATSKS-----KQVSLGLDSSLNFKWELSIG-----GQTLISKTEFNR 477
ISLTAEVKSK-----KQQR-LSLQSLLSYKLNLAIG-----DKTISKKDFEK 473
VEVSATLPSD-----RPS--VGLEALLQFRWELSLG-----GQRLTKAEVER 436
VEVSATLPSD-----RPS--VGLEALLQFRWELSLG-----GQRLTKAEVER 436
LKIIATLPPP-----ATNG-LTIDSLMQFQWQLQLG-----QHPLSEADFDQ 429
LKLEANAPGR-----NGSG-LGMQSLLLAFKWELSLA-----GKTLSRAEFFDR 433
LAIQAELPES-----SLG--VMLGESLNWDWELMIG-----GINLSMKELEM 487
IAIKAELATS-----ARG--LTLRENLEWSWELMIG-----GSILSLKDLEQ 494
LAIKAELSER-----SRG--FTLGETLDWSWELMIG-----GVTLLTLELER 527
LAIKAELSER-----SRG--FTLGETLDWSWELMIG-----GVTLLTLELER 527
LAIKAELSER-----SRG--FTLGENLDWSWELMIG-----GVTLLTLELER 490
LAIKAELPKR-----SRG--FTLGENLDWNWELMIG-----GVTLLTLELER 493
LAIQAELPEK-----SRG--FTLGETLDWSWELMIG-----GVTLLTLELER 488
LSIKAELPER-----SSG--FTLGECLAWENDLMIG-----GVTLLTLELER 491
LAIKAELPER-----SSG--FSLGESLDWSWDLMIG-----GVTLLTLELER 491
LAIKADLPDR-----SSG--FTLGESLDWSWDLMIG-----GVTLLTLELER 491
LSIEADLPER-----SRG--FSLGESLQSWELMIG-----GVTLLTRDLER 466
LVLSAYTPVDGVV-GKASKFGREQLVEFRWELAVG-----DDPLSEEEIAA 439
LVLSAYTPVDGVV-GKASKFGREQLVEFRWELAVG-----DDPLSEEEIAA 439
LAVSSAAPAA-----ESTVGMQGLLSYRWELAVG-----DKVLTRAEMER 389
MRVGASTKAAGAV-GGTAGLGLDALLRVDWDAVLG-----DQPLSAQELAL 499
MRLSP-----AG-GSPSMFGLHQIVNVRWDVALG-----GTPLTLDEL RH 436
AQANVKGKK-LKA-GYG--LTLDKIVSFDWEIALG-----DRALTVRELQA 489
AQANVKGKKKLQA-GYG--LTLDEIVSFDWEIALG-----DRVLTVRELQA 496
IKTKVKLPSMKGS-GSG--LTLDRMVACDYAAAALG-----NEELDLQELKT 438
ARLRFKPAEGKA-GKSQ-FTMDTLVSYDWRALALG-----DQEITETEF RK 470
LKASSGRKVAATVGESKSYFGLDSLQYQYELAIG-----EQTLTPQEW EQ 488
VOLKQ NATQT-----QSFFGMNTLVNFDWRI STN-----GIDLSESEFFE 351
HARRRDPPTH-----QTRIGLQELLSFDYRIAIG-----DESFSPEFFWE 478

.

:

:

:

FIGURE 8 (continued)

Synco_SNF2	LLAQKSP	LVVL	GEWITLQ	PADVRAAKVILQ	QQQS-APPLT	VEDALRLSI	511
Anava_SNF2	LVALKSP	LVEINGEW	VELRPQDIKTA	EAFFAARKD-Q	MALSLEDALRLSS		545
Nostoc_SNF2	LVALKSP	LVEINGEW	VELRPQDIKTA	EAFFTARKD-Q	MALSLEDALRLSS		542
Nodsp_SNF2	LVALNSP	LVEINGEW	VELRPQDIKTAQ	TFASRKD-EMT	LSLEDALRLSS		562
Lyn_sp_SNF2	LVAQESP	LVEINGEW	VELRPTDIKAAK	AFSSSRKD-Q	LSLTLEDALRLST		526
Crowa_SNF2	LLAQKSP	LVEVKGEWIALQ	PADVKAQQI	LNKSYD-PL	ELSVEDALRFST		522
Synel_PCC6301_SNF2	LAALET	PLVEINGDWIEVR	PQDIESAREFFR	KRKD-QPNLT	LADAIAS		485
Synel_PCC7942_SNF2	LAALET	PLVEINGDWIEVR	PQDIESAREFFR	KRKD-QPNLT	LADAIAS		485
Theel_BP-1_SNF2	LRRQGT	PLVYLN	GEWVLLRPQEV	KAAQEFLOS-PP	KTQLSLAETLRIAT		477
Glovi_SNF2	LAASEP	LVKVN	DNWVELRPQDV	RAAHSFLQSRKD-Q	VGLSLEDVLRNLF		482
Proma_CCOMP1375_SNF2	LAKKNSP	LLNHKG	TWIELRPNDL	KNASKFFAN---	TPELNLDKALRLSA		533
Proma_MIT\9211_SNF2	LASKRSP	LVRYKDS	WLELRPNDL	KIAEKFCSN---	NPELSDDALRLTA		540
Proma_MIT\9303_SNF2	LASKRSP	LVNHKG	AWIELRPNDL	KNAEHFCV---	NPGISLDDALRLTA		573
Proma_MIT9313_SNF2	LASKRSP	LVNHKG	AWIELRPNDL	KHAEHFCV---	NPGISLDDALRLTA		573
Syn_sp_CC9311_SNF2	LAGKRSP	LVRHKG	AWIELRPNDL	KNAERFCAA---	NPDLSLDDALRLTA		536
Syn_sp_WH\7805_SNF2	LAGKRSP	LVRHKG	AWIELRPNDL	KNAERFCAA---	NPDLSLDDALRLTA		539
Syn_sp_RS9916_SNF2	LAKKRSP	LVRHKG	TWIELRPNDL	KNAERFFAA---	KPDLSLDDALRLTA		534
Syn_sp_CC9605_SNF2	LSGKRSP	LVRHKG	AWIELRPNDL	KNAERFCGA---	KPELSLDDALRLTG		537
Syn_sp_WH\8102_SNF2	LSGKRSP	LVRHKG	AWIELRPNDL	RNAERFCGA---	NPELSLDDALRITA		537
Syn_sp_CC9902_SNF2	LSGKRSP	LVRHKG	AWIELRPNDL	RNAERFCGA---	NPELSLDDALRLTA		537
Syn_sp_WH\5701_SNF2	LAKKRSP	LVQHK	AWIELRPGDL	RNAEKFCAL---	DPVLSLDDALRLTG		512
Mycu_SNF2	LTETKSP	LIRLRGQ	WVALDTEQ	MRRGLEFLERKP--	TGRKTTAEIL-ALA		486
Mycbo_SNF2	LTETKSP	LIRLRGQ	WVALDTEQ	LRRGLEFLERKP--	TGRKTTAEIL-ALA		486
Nocfa_IFM\10152_SNF2	LVRAKSD	LVQLRG	EWQADHKVLA	AAARYVA AHLD-TSP	VTLADLLGEIA		438
Myxxa_DK_SNF2	LAQRKAP	LVFRGEW	VADPLELDAI	QRHLAQGP	G-RMALSEAVRVSLG		548
Synth_IAM14863_SNF2	LARQKR	PLVQMQGR	WVRVDERTL	AAVLRRIEQHG	G-QMELGTALRLAPEA		485
Metac_C2A_SNF2	LAKLKAP	LVKFRGQ	WVEVNDAEI	RAALEFWKKNP--	HGEASLREVLKLAV		537
Metma_Go1_SNF2	LAKLKAP	LVKFRGQ	WVEVNDAEI	RAALEFWKKNP--	NGEASLREVLKLAV		544
Pelph_BU-1_SNF2	LANLKV	PLVVRGQ	WTQIDHKEL	ANALHFLEKHP--	TGELSARELLSTAL		486
Archaeon_RC-I_SNF2	LAAALKEP	LLQIGK	WFALKKEDID	SIMKA	FRACK--TGEMALSEALRLNG		518
Nos_sp_PCC7120_SNF2	LINTKAP	LVHFRGQ	WMELDRDKMQ	QLLEFFWQSHG	DEQPQMSLLEFMQ	RSA	538
Bacce_ATCC10987_SNF2	LVEQNKR	LFNINGQ	WMRLDPAFIE	EVRLKLMNRAD-KYG	-LEMKDVLQQHL		399
Methu_JF-1_SNF2	KVKEKAP	FIWLGNRWIS	FHPDAIQ	HALDSFSRHQ-SKGG	DTIGDLLRLSL		527

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myctu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon\RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

555
589
586
606
570
566
529
529
521
526
577
584
617
617
580
583
578
581
581
581
556
531
531
475
588
518
581
588
530
561
576
449
569

GD LQT VSKLP-----VTQFAARGILQELIDTLRNPEGVKAIADPPGFQ
GDTQVIEKLP-----VVSFEASGALQELIGALTNNQAVAPLPTPKNFQ
GDTQVIEKLP-----VVSFEASGALQELIGALTNNQAVAPLPTPKNFQ
GDTQAIKLP-----VVSFEASGTLQELIGALTNNQAI SPLTPANFQ
GDSQMVEKLP-----IVNFEAGKLEELLNTLTNNRSLDEIKTPSNFQ
GDISTVAKLP-----ITNFEAKGELANLINAINNNEIPMIENPRGFKG
GESPNVGRLP-----VVNFEAAGLLEEALAVFQGQSRSPAALPAPPTFQ
GESPNVGRLP-----VVNFEAAGLLEEALAVFQGQSRSPAALPAPPTFQ
GDTVTVAKLP-----ILGLDTNDALQTLLDGLTGKQSLDPVPTPQEFQ
GDTPKIDGLP-----IVNFDSSGPIQQLLETLTLDQRKLTPIDEPPGFKG
NKGNTFMKLP-----VHHFESGPRLSVLEQYHHQKAPEPLPAPNGFHG
TKGETLMKLP-----VHQFNAGPKLQGVLEQYHQHTSPEPLAAPDGFY
TDGDTLMRLP-----VHRFEAGPRLQAVLEQYHQKAPDPLPAPEGFCG
TDGDTLMRLP-----VHRFEAGPRLQAVLEQYHQKAPDPLPAPEGFCG
TEGDTMMRLP-----VHQFDAGPRLQAVLEQYHQKAPDPLPAPEGFSG
SEGD TLMRLP-----VHAFDAGPRLQGVLEQYHQKAPDPLPAPEGFCG
SEGD TLMRMP-----VHRLEAGPRLQAVLEQYHQKAPDPLPAPEGFCG
TEGELLMRMP-----VHRFDAGPRLQSVLQQYHQKAPDPLPAPEGFSG
TEGDLLMRLP-----VHRFEAGPRLQAVLEQYHQKAPDPLPAPEGFCG
TEGELMMRLP-----VHRFDAGPRLQGVLEQYHQKAPDPLPAPEGFSG
NEGETLQRLP-----VHRFTAGPRLKAVLEQYHQKAPDPLPAPEGFAG
ASHPDDVDTPLE----VTAVRADGWIGDLLAGAAA-ASLQPLDPPDGFTA
ASHPDDVDTPLE----VTAVRADGWIGDLLAGAAA-ASLQPLDPPDGFTA
ATRVDKVP-----LTEVTATGWAGELFDGGR-----EPVATPGGLKA
ETRHGQLP-----VTVLATGALEERLRLRE-GGATAQDAPRALRA
DEAT-----ATGWIAELLERLQEPARMEPVPTPGGFAG
GVSEKADGVD-----VEGLNAAGWIEELIRRLKDKTGFEELPAPDGFSG
GVSEKADGVN-----VEGLNATGWIGELISRLKDKTGFEELPAPNGFSG
GAQKKEDALF-----LRSVEIEGWIQELLEKLSQGQFELLPPPEHFEG
GLEDFN-GIP-----VSGMKSSGWLAE LFDRLAAGEKITSLAPPDG FNG
QGEDD-----WEIEYDAALSEIMAKLQDKSQLEPISEDNLNQ
SNTAETEIVEEDSPFTDIEIELDGYEDLFQKLLHIGDIPKVDVPSSLNA
KKMEDSAVP-----VSIHAKDDWVADLLDFFRTRTETNQAVVPKKFKG

:

:

FIGURE 8 (continued)

	Sequence	Position
Ssynco_SNF2	TLRPYQARGVGWLAFLERWGLGAC	605
Anava_SNF2	KLRPYQERGAAWLAFLERWGLGAC	639
Nostoc_SNF2	QLRPYQERGAAWLAFLERWGLGAC	636
Nodsp_SNF2	QLRPYQERGAAWLAFLERWGLGAC	656
Lyn_sp_SNF2	ELRPYQARGVSWLAFLERWGLGAC	620
Crowa_SNF2	QLRPYQQRGVGWSLFLERWGLGAC	616
Synel_PCC6301_SNF2	ELRPYQERGVGWSLFLERWGLGAC	579
Ssynel_PCC7942_SNF2	ELRPYQERGVGWSLFLERWGLGAC	579
Theel_BP-1_SNF2	ELRPYQARGVAWLSFLERWGLGAC	571
Glovi_SNF2	TLRPYQKIGVGWLAFLERWGLGAC	576
Proma_CCMF1375_SNF2	QLRPYQERGLGWLAFYRFKQGAC	627
Proma_MIT\9211_SNF2	QLRPYQERGIWLAFLHRFNQGAC	634
Proma_MIT\9303_SNF2	QLRPYQERGLGWLAFHRFDQGAC	667
Proma_MIT9313_SNF2	QLRPYQERGLGWLAFHRFDQGAC	667
Syn_sp_CC9311_SNF2	QLRPYQERGLGWLAFHRFDQGAC	630
Ssyn_sp_WH\7805_SNF2	QLRPYQERGLGWLAFHRFDQGAC	633
Ssyn_sp_RS9916_SNF2	QLRPYQERGLGWLAFHRFDQGAC	628
Ssyn_sp_CC9605_SNF2	QLRPYQERGLGWLAFHRFDQGAC	631
Ssyn_sp_WH\8102_SNF2	QLRPYQERGLGWLAFHRFDQGAC	631
Ssyn_sp_CC9902_SNF2	QLRPYQERGLGWLAFHRFDQGAC	631
Ssyn_sp_WH\5701_SNF2	QLRPYQERGLGWLAFHRFDQGAC	606
Myctu_SNF2	TLRPYQQRGLAWLAFLSSGLGSC	581
Mycbo_SNF2	TLRPYQQRGLAWLAFLSSGLGSC	581
Nocfa_IFM\10152_SNF2	QLRPYQLRGLSWLATMSRMCGGI	525
Myxxa_DK_SNF2	TLRPYQSRGLHWLDTLASGLGAC	638
Synth_IAM14863_SNF2	TLRPYQQRGLAWLAFLRRWGLGAC	567
Metac_C2A_SNF2	TLRPYQFRGYSWLAFLRQWIGIGAC	631
Metma_Go1_SNF2	TLRPYQFRGYSWLAFLRQWIGIGAC	638
Pelph_BU-1_SNF2	TLRLYQERGFWSLFLRKWGLGAC	576
Archaeon_RC-I_SNF2	ELRDYQVKGYSWLAFLMKKYGLSI	607
Nos_sp_PCC7120_SNF2\II	NLREYQKRGVAWLQYLEKLGNGCL	626
Bacce_ATCC10987_SNF2	TLRPYQQHGI EWLLYLRLKLGFGAL	499
Methu_JF-1_SNF2	ILRPYQEEGFSLCQCTRRGFGAC	619

[illegible]

KTLHHGDRRK-KGQPLVKQVKDQQIVLTSYALLQDRFSSLKLVDWQGIV 681
KVLQYHGDKRP-KGKAFFEAVKNHDLVITSYSLIHRDIKSLQGLSWQIIV 715
KVLQYHGDKRP-KGKAFFEAVKKHDLVITSYSLIHRDIKSLQGIPWQIIV 712
KVLQHHGDKRL-KGKAFFEAVKKHDVIIITSYSLVHRDIKSLQSVDWQTVV 732
KVTVHHGDKRQ-KGKNFAQQAQKYNLIITSYPLTFRDEKELKTVNWKGLV 696
STLIHHGDKRS-KGKAFFVKA VSKKNV IITSYSLIYRDIKSFEQVEWQGIV 692
RWKLHYGPDRA-QGKALATALKDCDLVLTYSYSLVARDQKAI AAI DWQGIV 655
RWKLHYGPDRA-QGKALATALKDCDLVLTYSYSLVARDQKAI AAI DWQGIV 655
RAYVHHGSDRP-KGKAFLKKVETHDLILT SYALLQDR T TLQQVLWQH LV 647
SVHVHHGARRP-KGRNFVETAQKKQIIVSSYALVQRDSKDLKRVEWLGLV 652
CIHEHYGSKRHSSIPKLQNYLKKVDIMITSYGLLYRDGELLQEIDWQGIV 704
SVLEHYGPNRSSTLLKKIILKKVDILITSYGLLHRDKQLLKTIDWQGI 711
NVREHYGPRRPSTPAALKKALKGLDLVLT SYGLLQRDSELLETVDWQGVV 744
NVREHYGPRRPSTPAALKKALKGLDLVLT SYGLLQRDSELLETVDWQGVV 744
KVHEHYGPKRPSTPAALKKALKDVLDVLT SYGLLQRDSELLESHDWQGLV 707
TVHEHYGPKRPSTPAALKKALKDVLDVLT SYGLLQRDSELLESEFWDWQGTV 710
EVKEHYGPRRPATPAALKKSLKDVLVLT SYGLLQRDSELLESLDWQGVV 705
KVTEHYGPRRPSTPAELKKALKEVDLVLT SYGLLQRDSELLETTQDWQGVV 708
AVREHYGPRRPSTPAALKKALKDVLDVLT SYGLLQRDSELLESDWQGVV 708
SVKEHYGPRRPSTPAALKKELKDVLVLT SYGLMQRDSELLESDVDWQGVV 708
NVVEHYGPRRPSTPAALKKKLEGMDLVLT SYGLLQRDSELLESSLDWQGVV 683
RVYAHHG GARLHGEALRDHLERT-DLVVSTYTTATRDI DE LAEYEWNRV 660
RVYAHHG GARLHGEALRDHLERT-DLVVSTYTTATRDI DE LSEYEWNRV 660
RVLVHHGADRRRDAELDAAVADS-DLVLT TYAILARDAAE LSRQSWDRV 602
RLTRHYGAERARAANRFP RAPGA--VVLT TYGLLRRDAALLARVDWGAVV 714
RVLVHHGPGRLGEPD-FARQAGAHDVVLT TYSSLARDAALLGQVTWNGIV 644
SVMVHHGTSRK-KEEEFKKEATNHSIVVSSYGLLQRDLKFLKGVSWAGVV 726
SVMVHHGTSRK-KEEEFKKEAMNHAI VISSYGLVQRDLKFLKEVHWAGVV 737
AVLVHHGIDRM-KTADFRKAASASALVISSYGLLQRDL EFLSKVPWAGII 655
KVHIHHGAGRA-DKEQFGKIVKAHDLILSTYAHAYRDEELLKEVNWKLVV 686
KTMVHHGSDRLQDAAEFKSAQQQHDVVISSFTLARLDEKLLNSVTWQRLV 703
RVQLHYGSNRA-KGEPFKDFLQ SADVVLT SYALAQ LDEEELSTLCWD A VI 575
RSWVHHGTDRC-KGDDFVRHVGSYDLVLT TYH LAARDVDHLKTVPWSAII 695
::* *
::: *
:

Synco_SNF2
Anava_SNF2
Nostoc_SNF2
Nodsp_SNF2
Lyn_sp_SNF2
Crowa_SNF2
Synel_PCC6301_SNF2
Synel_PCC7942_SNF2
Theel_BP-1_SNF2
Glovi_SNF2
Proma_CCMP1375_SNF2
Proma_MIT\9211_SNF2
Proma_MIT\9303_SNF2
Proma_MIT9313_SNF2
Syn_sp_CC9311_SNF2
Syn_sp_WH\7805_SNF2
Syn_sp_RS9916_SNF2
Syn_sp_CC9605_SNF2
Syn_sp_WH\8102_SNF2
Syn_sp_CC9902_SNF2
Syn_sp_WH\5701_SNF2
Myc tu_SNF2
Mycbo_SNF2
Nocfa_IFM\10152_SNF2
Myxxa_DK_SNF2
Symth_IAM14863_SNF2
Metac_C2A_SNF2
Metma_Go1_SNF2
Pelph_BU-1_SNF2
Archaeon\RC-I_SNF2
Nos_sp_PCC7120_SNF2\II
Bacce_ATCC10987_SNF2
Methu_JF-1_SNF2

FIGURE 8 (continued)

FIGURE 8 (continued)

Synco_SNF2	ILEFLNPGFLGNQSFQRRFANPIEKFGDRQSLILRNLVRPFILRRLKT	774
Anava_SNF2	ILDFLNPGYLGKQFFQRRFAMPIEKYGDAASLNQLRALVQPFILRRLKT	808
Nostoc_SNF2	ILDFLNPGYLGKQFFQRRFAMPIEKYGDAASLNQLRALVQPFILRRLKT	805
Nodsp_SNF2	ILDFLNPGYLGNRQFFQRRFAMPIEKYGD TASLNQLRGLVQPFILRRLKT	825
Lyn_sp_SNF2	IMDFLNPGLGQRQFFQRRFAPIEKYGD TDSLKTLRSLVQPFILRRLKT	789
Crowa_SNF2	ILDFLNPGFLGTQFFRRRFATPIEKYGD KESLQIMRSLVRPFILRRLKT	785
Synel_PCC6301_SNF2	IVEFLQPGHLGTPFFQKRFVTPIERFGDADSLTALRQRVQPLILRRLKT	754
Synel_PCC7942_SNF2	IVEFLQPGHLGTPFFQKRFVTPIERFGDADSLTALRQRVQPLILRRLKT	754
Theel_BP-1_SNF2	IMDFLHPGYLGHRTYFQHRYVRPIERYGD TTSNALRTYVQPFILRRLKT	740
Glovi_SNF2	ILDFLNPGYLGARNFFQRRFAVPIEKYGD RSSANALKALVQPFILRRLS	745
Proma_CCMP1375_SNF2	LMDFLNP KVLGEEDFFNQRYKLPIEHYGD ISSLKD LKTQVSPFILRRLKT	802
Proma_MIT\9211_SNF2	LMDFLNP SVLGEKEFFDQRYKLPIERYGD ISSLTD LKARVSPFILRRLS	809
Proma_MIT\9303_SNF2	LMDFLNP RVLGEEDFFRQRYRLPIERYGD MSSL RDLKGRVGFILRRLKT	842
Proma_MIT9313_SNF2	LMDFLNP RVLGEEDFFRQRYRLPIERYGD MSSL RDLKGRVGFILRRLKT	842
Syn_sp_CC9311_SNF2	LMDFLNP RVLGEEEEFFRHRYRMP IERYGD LSSL RDLKARVGFILRRLKT	805
Syn_sp_WH\7805_SNF2	LMDFLNP NVLGEEEFFRQRYRMP IERYGD MSSL RDLKSRVGFILRRLKT	808
Syn_sp_RS9916_SNF2	LMDFLNP RVLGEEDFFRQRYRMP IERYGD MSSL RDLKSRVGFILRRLKT	803
Syn_sp_CC9605_SNF2	LMDFLNP KVLGEEDFFRQRYRMP IERYGD MSSL RDLKGRVGFILRRLKT	808
Syn_sp_WH\8102_SNF2	LMDFLSP KVLGEEDFFRQRYRMP IERYGD MASL RDLKARVGFILRRLKT	806
Syn_sp_CC9902_SNF2	LMDFLNP KVLGEEDFFRQRYRMP IERYGD MSSL RDLKARVGFILRRLKT	806
Syn_sp_WH\5701_SNF2	LMDFLNP KVLGEEEEFFRQRYRLPIERYGD MASVRDLKARVGFILRRLKT	781
Mycu_SNF2	IMDFLNP GLLGSSSERFRTRYAIP IERHGHTEPAERLRASTRPYILRRLKT	753
Mycbo_SNF2	IMDFLNP GLLGSSSERFRTRYAIP IERHGHTEPAERLRASTRPYILRRLKT	753
Nocfa_IFM\10152_SNF2	IMDFAVP KLLGTAPTFRARFAVPIERGQDPNALSR LRF LTQPFVLRVKA	695
Myxxa_DK_SNF2	ILEFANP GLLGPLETFRRELALPIERHGNQEASARLRRLVSPFVLRRLKS	807
Symth_IAM14863_SNF2	LFQFLNPGLLGSREEFERRYAVPIQRYQDEEAAARLR RQVGFILRRQKN	737
Metac_C2A_SNF2	IMEFLNPGFLGNQAGFKRNFFIPIQ AERDQEAARRLKEITGPFILRRLKT	819
Metma_Go1_SNF2	IMEFLNPGFLGSQAGFKRNFFIPIQ AERDQEAARRLKEITGPFILRRLKT	830
Pelph_BU-1_SNF2	LMDFLNP GFLGTQHFFKQNFYTP IQWYGDPEASARLKS LTGPFILRRMKS	748
Archaeon_RC-I_SNF2	IVDFLNP GYLGAETFRKQFAIP IERYDDAARSEK LKQAIKPLVLRVKT	779
Nos_sp_PCC7120_SNF2\II	IFNFLNPGYLGKEAQFRKSFEIPIQKDN DKVSTTLKKLVEPLILRRVKT	796
Bacce_ATCC10987_SNF2	IFDFINHG YLGS LGQFQRRFVSPIEKDRDEGKIQQVQRFISPFILRRTKK	668
Methu_JF-1_SNF2	IMDFLNP GYLGSQSAFTNRYSRPIEQEKNT ELIQELRSLIRPFILRRMKT	788
	:.: * ** * ** :	*** *

FIGURE 8 (continued)

Synco_SNF2	DQTI IQDLPEKQEMTVFCDL SQEQAGLYQQLV EESLQAIADSEG-IQRRHG	823
Anava_SNF2	DRDIIQDL PDKQEMTVFCGLTGEQAALYQKV VETSLAEIESAEG-LQRRG	857
Nostoc_SNF2	DRDIIQDL PDKQEMTVFCGLTGEQAALYQKAVETSLAEIESAEG-LQRRG	854
Nodsp_SNF2	DRDIIQDLPEKQEMTVFCGLAAEQAAALYQQVVEASLVEIESAEG-LQRRG	874
Lyn_sp_SNF2	DREIIQDLPEKQENTIFC SLSTEQATLYQKIVDQSLADIDS AAG-IQRRG	838
Crowa_SNF2	DKTIIQDLPEKQEMTIFCGLSSEQKLYQQLV DNSLVAIEEKTG-IERKG	834
Synel_PCC6301_SNF2	DRSIIADLPEKQEMTVFCPLVQEQAADRYQVLVNEALANIEASEG-IQRRG	803
Synel_PCC7942_SNF2	DRSIIADLPEKQEMTVFCPLVQEQAADRYQVLVNEALANIEASEG-IQRRG	803
Theel_BP-1_SNF2	DRSIIQDLPEKQEMLVYCGLTLEQMQLYTAVVEDSLAAIENSQG-IQRRG	789
Glovi_SNF2	DPQIIQDLPEKQETNVFCPLTPEQAALYERVVNESLAKIEQSTG-IQRRG	794
Proma_CCMP1375_SNF2	DQSIISDL POKIELNEWVGLSQEQELL YKQTV EKSLDELASLPI-GQRQG	851
Proma_MIT\9211_SNF2	DKSIIISDLPSKVELKEWITLSQEQRALYNKTV DNTLQEIARSPI-GQRHA	858
Proma_MIT\9303_SNF2	DKAIIISDLPEKVELSEWVGLSKEQAALYRNTVDETLEA IARAPS-GQRHG	891
Proma_MIT9313_SNF2	DKAIIISDLPEKVELSEWVGLSKEQAALYRNTVDETLEA IARAPR-GQRHG	891
Syn_sp_CC9311_SNF2	DKAIIISDLPEKVELSEWVGLSKEQKSLYAKTVEDTLDAIARAPR-GKRRHG	854
Syn_sp_WH\7805_SNF2	DKAIIISDLPEKVELSEWVGLSKEQKSLYAKTVENTLDAIARAPR-GKRRHG	857
Syn_sp_RS9916_SNF2	DKAIIISDLPEKVELSEWVGLSREQKALYAKTVEDTLDAIARAPR-GQRHG	852
Syn_sp_CC9605_SNF2	DKTIIISDLPEKVELSEWVGLSKEQKSLYSKTVEDTLDAIARAPR-GQRHG	857
Syn_sp_WH\8102_SNF2	DKTIIISDLPEKVELSEWVGLSKEQKSLYSKTVEDTLDAIARAPR-GQRHG	855
Syn_sp_CC9902_SNF2	DKSIIISDLPEKVELSEWVGLSKEQKSLYNKTVEDTLDAIATAPR-GQRHG	855
Syn_sp_WH\5701_SNF2	DRSIIISDLPEKVELKEWVGLSPEQVKLYRRRTVEDTLDAIARAPV-GQKHG	830
Myctu_SNF2	DPAIISDLPEKIEIKQYQCQLTTEQASLYQAVVADMM EKIENTEG-IERRG	802
Mycbo_SNF2	DPAIISDLPEKIEIKQYQCQLTTEQASLYQAVVADMM EKIENTEG-IERRG	802
Nocfa_IFM\10152_SNF2	DPAVIGDLPDKLEMTVRANLTVEQAALYQAVVDDMLVKLRS AKG-MARKG	744
Myxxa_DK_SNF2	DPTIITDLP AKNEMKVVCTLTREQASLYKAVVDEELRRRIEEADG-MERRG	856
Symth_IAM14863_SNF2	DPAIAPDLPDKLENTELVTL SVEQAALYEAIVQETLERAAQADG-IQRQA	786
Metac_C2A_SNF2	DTSIIISDLPEKMEMKTYCTLTKEQASLYAAAVLEDI EETMEEAEEGIQRK	869
Metma_Go1_SNF2	DTSIIISDLPEKMEMKTYCTLTKEQASLYAAAVLEDI REAIEGAEEGIQRK	880
Pelph_BU-1_SNF2	DKSIIISDL PDKIEMKEYCSLTKEQASLYKAVVDELQEKIESAEG-IDRRG	797
Archaeon_RC-I_SNF2	DPAI IKDLPDKIEIKEPCNLTKEQATLYEAIVENMLKSIDKATA-MQRRG	828
Nos_sp_PCC7120_SNF2\II	DQSI IKDLPDKVEQKLYTNLTKEQASLYEVVVRDV EEKLEQAEAG-IQRKG	845
Bacce_ATCC10987_SNF2	DQTV ALNLPDKQEQA KAYCPLTGEQASLYEQLVQD TLQNV EGLSG-IERRG	717
Methu_JF-1_SNF2	DKHVIDDLP EKMENRVYCTLTPEQATLYQAVVLDMAKNL DKVEG-IARKG	837
	* : ; ** * * * * * * * * * :	

FIGURE 8 (continued)

Synco_SNF2	LVLTLTKLKQVCNHPDLL-----LKKPAITHGH--QSGKLIRL	860
Anava_SNF2	MILALLIKLKQICNHPAQY-----LKTNTLEQY---SSGKLQRL	893
Nostoc_SNF2	MILALLIKLKQICNHPAQY-----LKINTLEQH---SSGKLQRL	890
Nodsp_SNF2	MILALLVKLKQICNHPAQY-----LKAATLQEH---SSAKLQRL	910
Lyn_sp_SNF2	MILALLVKLKQVCNHPILLNGKATKTGKKKQVETQGLSLQ---SSGKLQRF	885
Crowa_SNF2	LILSLLLKLKQICNHPAHF-----LKQKSLKTAE--QSGKLLRL	871
Synel_PCC6301_SNF2	QILALLTRLKQLCNHPSLL-----LEKPKLDPNFGDRSAKLQRL	842
Synel_PCC7942_SNF2	QILALLTRLKQLCNHPSLL-----LEKPKLDPNFGDRSAKLQRL	842
Theel_BP-1_SNF2	NILATLTKLKQICNHPAQY-----LKQEDYAP---DRSGKLQRL	825
Glovi_SNF2	TVLATLVKLKQICNHPSHY-----LGDDGPLAN---RSGKLSRL	830
Proma_CCMP1375_SNF2	KTGLLTRLKQICNHPAIA-----LKETQVEKNFLLRSSKLQRL	890
Proma_MIT\9211_SNF2	KTGLLTRLKQICNHPALA-----LKEKNISDDFGIRSTKLQRL	897
Proma_MIT\9303_SNF2	KVLGLLTRLKQICNHPALA-----LKEKTVAKGFMDRSAKLLRL	930
Proma_MIT9313_SNF2	KVLGLLTRLKQICNHPALA-----LKEQTVAKGFMDRSAKLLRL	930
Syn_sp_CC9311_SNF2	QVLGLLTKLKQICNHPALA-----LKEQGASEDFLKRSVKLQRL	893
Syn_sp_WH\7805_SNF2	QVLGLLTRLKQICNHPALA-----LKEEVAGDDFLQRSVKLQRL	896
Syn_sp_RS9916_SNF2	QVLGLLTKLKQICNHPALA-----LKEEAAGDEFLQRSMKLQRL	891
Syn_sp_CC9605_SNF2	QVLALLTRLKQICNHPALA-----LSEGAVDDGFLGRSAKLQRL	896
Syn_sp_WH\8102_SNF2	QVLGLLTRLKQICNHPALA-----LSENAVDDGFLGRSAKLQRL	894
Syn_sp_CC9902_SNF2	QVLALLTRLKQICNHPALA-----QREGAVDSEFLGRSAKLMRL	894
Syn_sp_WH\5701_SNF2	QVLGLLTKLKQVCNHPALM-----LKEGEVGAGFSARSACLQRL	869
Myctu_SNF2	NVLAAMAKLKQVCNHPAQL-----LHDRSPVGRR---SGKVIRL	838
Mycbo_SNF2	NVLAAMAKLKQVCNHPAQL-----LHDRSPVGRR---SGKVIRL	838
Nocfa_IFM\10152_SNF2	AVLGALTRLKQVCNHPAHF-----LGDGSPVLHRGRHRSGLALV	784
Myxxa_DK_SNF2	RVLALLLYTKQIANHPAQY-----LGESGPLPGR---SGKLARV	892
Symth_IAM14863_SNF2	AVLAGLTRLKQVCNHPAAA-----TGD-GPLVGR---SGKIDRL	821
Metac_C2A_SNF2	IILSALTRLKQVCNHPAQFLK-----DNSAVPG---RSGKLARL	905
Metma_Go1_SNF2	IILSALSRLKQVCNHPAQFLK-----DNSTIPG---RSGKLARL	916
Pelph_BU-1_SNF2	LVLALLVKLKQVCNHPAHLLG-----DNSAIAH---RSGKIKRL	833
Archaeon_RC-I_SNF2	IVLASLMKLKQVCDHPSLYIK-----TGAVTDDKTILRSGLKRL	868
Nos_sp_PCC7120_SNF2\II	LILSTLMKLKQICNHPRQFLQ-----DNSEFLPE---RSHKLSRL	882
Bacce_ATCC10987_SNF2	FILLMLNKLKQICNHPALYL-----KETEPKDI IERSMKTSTL	755
Methu_JF-1_SNF2	AILAAITRLKQICNHPPGRVG-----RDKTIK--AERSGKVSRL	873
	* ; **:::** * *	

FIGURE 8 (continued)

Synco_SNF2	AEMLEEIISEG-----DRV	GIFTQFASWGHLKPYLEKYFN---QEV	899
Anava_SNF2	EEMLEEVLAESNTYGVAGAGRA	LIFTQFAEWGKLLKPHLEKQLG---REV	940
Nostoc_SNF2	EEMLEEVLAESNTYGVAGAGRA	LIFTQFAEWGKLLKPHLEKQLG---REI	937
Nodsp_SNF2	DEMLTVALEEG-----DRA	LIFTQFAEWGKLLKAHLQQTLG---KEI	949
Lyn_sp_SNF2	KEMLEELLSEG-----DRA	IVFTQFAEWGKVLQPYLEQQLN---REV	924
Crowa_SNF2	EEMLEELIEEG-----DHA	LIFTQFSEWGKLLQPYLQKKFQ---QDV	910
Synel_PCC6301_SNF2	LEMLAELTDAG-----DRA	LVFTQFAGWGSLLQQFLQEQLG---REV	881
Synel_PCC7942_SNF2	LEMLAELTDAG-----DRA	LVFTQFAGWGSLLQQFLQEQLG---REV	881
Theel_BP-1_SNF2	IEMLQALQEVG-----DRA	LVFTQFAEFGTHLKTYLEKALQ---QEV	864
Glovi_SNF2	GEMLEEVLADE-----ERA	LIFTQFAEWGHLLQAHLSRQLG---SEV	869
Proma_CCMP1375_SNF2	EEILQEVKESH-----DRA	LLFTQFAEWGHLLQAYLQTKWE---SEV	929
Proma_MIT\9211_SNF2	EELLDVIFATE-----DRA	LLFTQFAEWGHLLQAYLEKKWG---HSI	936
Proma_MIT\9303_SNF2	EEILEEVIEAG-----DRA	LLFTQFAEWGHLLKAYLQQRWR---FEV	969
Proma_MIT9313_SNF2	EEILEEVIEAG-----DRA	LLFTQFAEWGHLLKAYLQQRWR---FEV	969
Syn_sp_CC9311_SNF2	EEILDEVVEAG-----DRA	LLFTQFAEWGKLLQDYLQRRWR---SEV	932
Syn_sp_WH\7805_SNF2	EEILEEVIAAG-----DRA	LLFTQFAEWGHLLQGYLQRRWR---SEV	935
Syn_sp_RS9916_SNF2	EEILEEVIDAG-----DRA	LLFTQFAEWGHLLQGYLQRRWR---SEV	930
Syn_sp_CC9605_SNF2	EEILDEVIEAG-----DRA	LLFTQFAEWGHLLRAWMQQRWK---SEV	935
Syn_sp_WH\8102_SNF2	EEILDEVIEAG-----DRA	LLFTQFAEWGHLLQSWMQQRWK---ADV	933
Syn_sp_CC9902_SNF2	EEILEEVIEAG-----DRA	LLFTQFAEWGHLLQAWMQQRWK---SEV	933
Syn_sp_WH\5701_SNF2	EEIVEEVIAAG-----DRA	LLFTQFAEWGHLLQTHLQQRFH---QEV	908
Myctu_SNF2	EEILEEIIAEG-----DRV	LCFTQFTEFAELLVPHLAARFGRAARDI	880
Mycbo_SNF2	EEILEEIIAEG-----DRV	LCFTQFTEFAELLVPHLAARFGRAARDI	880
Nocfa_IFM\10152_SNF2	EDVLDTVVADG-----EKA	LLFTQFERFGDLLAPYLSERFG---API	823
Myxxa_DK_SNF2	VEMLEESLAAG-----DKA	LVFTQFERMGDKLVAHLSEYLG---HEV	931
Symth_IAM14863_SNF2	VQLLQEVLAAG-----EQA	LLFTQFARFGGRIQAYLAETLG---CEV	860
Metac_C2A_SNF2	TEMLDVILENG-----EKA	LVFTQFAEMGKMLKEHLQASFG---CEV	944
Metma_Go1_SNF2	TEMLDVVLENG-----EKA	LVFTQFAEMGKMVKEHLQASFG---CEV	955
Pelph_BU-1_SNF2	TELLGDIREAG-----EKT	LLFTQFTMMGMTMLQHYLQELYG---EEV	872
Archaeon_RC-I_SNF2	TELLEEALAEG-----DSV	LIFTQFVEMGEMLKAYLQSTFD---EEA	907
Nos_sp_PCC7120_SNF2\II	VEMVDEAISEG-----ESL	LIFSQFTEVCEQIEKYLKHNHLH---CNT	921
Bacce_ATCC10987_SNF2	MELIENIKDQN-----ESC	LIFTQYIGMGNMLKDVLEEHEFG---QRV	794
Methu_JF-1_SNF2	LEMIEEITSEG-----DSA	LIFSQYATFAEEELAGMIEKQGD---TPV	912
	:::	: * : *	:
			Motif IV

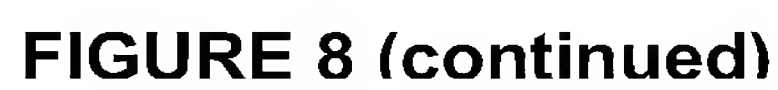
FIGURE 8(continued)

Ssynco_SNF2	LFLHGGTPAEQRQALVERFQ-QDPNSPYLFIILSLKAGGTGLNLTTRANHV	F948
Anava_SNF2	FFLYGSTSKKQREEMIDRFQ-HDPQGPPIMILSLKAGGVGLNLTTRANHV	F989
Nostoc_SNF2	FFLYGTSKKQREEMIDRFQ-HDPQGPPIMILSLKAGGVGLNLTTRANHV	F986
Nodsp_SNF2	FFLYGSSKKQREEMIDRFQ-HDPQGPPIMILSLKAGGVGLNLTTRANHV	F998
Llyn_sp_SNF2	LFLYGATRNKNKREEMIDRFQ-QDPQGPPIFILSLKAGGVGLNLTTRANHV	F973
Crowa_SNF2	LFLYGATRVRVQRQEMIDRFQ-QDPNGPRIFILSLKAGGTGLNLTTRANHV	F959
Synel_PCC6301_SNF2	LFLSGSTKKGDRQQMVDRFQ-NDPQAPAIFILSLKAGGVGLNLTKANHV	F930
Ssynel_PCC7942_SNF2	LFLSGSTKKGDRQQMVDRFQ-NDPQAPAIFILSLKAGGVGLNLTKANHV	F930
Theel_BP-1_SNF2	FFLSGRTPKAQRELWVERFQ-HDPEAPRVFILLKAGGVGLNLTTRANHV	F913
Glovi_SNF2	FFLYGTSKNQREAMIERFQ-SDPQGPRIFILSLKAGGVGLNLTTRANHV	F918
Proma_CMP1375_SNF2	PFLHGGTPKKRQEMIDRFQ-DDPRGPNIFFLLSLKAGGVGLNLTTRANHV	F978
Proma_MIT\9211_SNF2	LFLHGGTRKIDRQSMVDQFQ-EDPRGPKLFLLSLKAGGIGLNLTTRANHV	J985
Proma_MIT\9303_SNF2	PFLHGSTSATERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F1018
Proma_MIT9313_SNF2	PFLHGSTSATERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F1018
Ssyn_sp_CC9311_SNF2	PFLSGSTSSEERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F981
Ssyn_sp_WH\7805_SNF2	PFLSGSTSKEGERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F984
Ssyn_sp_RS9916_SNF2	PFLNGSTSSEERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F979
Ssyn_sp_CC9605_SNF2	PFLHGGTRKNERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F984
Ssyn_sp_WH\8102_SNF2	PFLHGGTRKNERQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F982
Ssyn_sp_CC9902_SNF2	PFLHGGTRKSDRQAMVDRFQ-EDPRGPQLFLLSLKAGGVGLNLTRASHV	F982
Ssyn_sp_WH\5701_SNF2	PFLYGSTSKEGERQAMVDRFQ-DDPRGPQLFLLSLKAGGVGLNLTRASHV	F957
Mycetu_SNF2	AYLHGGTTPRRRDENVARFQ-SGDGPP-IFLLSLKAGGTGLNLTAANHVV	V928
Mycbo_SNF2	AYLHGGTTPRRRDENVARFQ-SGDGPP-IFLLSLKAGGTGLNLTAANHVV	V928
Nocfa_IFM\10152_SNF2	PFLHGGVTCKNRDTMVERFQ-SGDGPP-VMLLSLKAGGTGLTLTAANHVV	V871
Myxxa_DK_SNF2	LFLHGGTTPRKARDENVRRFQ-EDVHGPRVFVLSVKAGGTGLNLTAASHV	F980
Ssymth_IAM14863_SNF2	LFLHGGTTPQPERDLVARFQ-AGEAP--LFILSLKAGGLGNLTAAATHV	F907
Metac_C2A_SNF2	LFLHGGVPRKQDRMLERFQEGKEYLP-IFVLISLKAGGTGLNLTGАНHV	F993
Metma_Go1_SNF2	LFLHGGVPRKQDRMLERFQEGKEYLP-IFVLISLKAGGTGLNLTGАНHV	F1004
Pelph_BU-1_SNF2	LFLHGGVTCKRRDEMVEVSFQKEEGSSPSIFILSLKAGGTGLNLTTАНHV	V922
Archaeon_RC-I_SNF2	LFLHGGVPQKARDKMVLRFGKED--GPRIFIVSLKAGGVGLNLTKASHV	F955
Nos_sp_PCC7120_SNF2\II	YYLHGGTSRQRREQMISDFQ-NPDTEASVFVLSLKAGGVGITLTKАНHV	F970
Bacce_ATCC10987_SNF2	LFLNGSVPKKERDKMIEQFQ--NG-TYDIFILSLKAGGTGLNLTAАНHV	I841
Methu_JF-1_SNF2	LLLTGSTPRKKREQMIEEFQ--ASTTPIIFVILSLKAGGTGLNLTKATHV	F960
	* * * : : *	: : : *
	* : * * * *	* : * * * *
		MOTIF V

[illegible]

Synco_SNF2	QLAEQTVDAGENWLTRLDTDKLRQLLTLSATPVDYQAEASD	1039
Anava_SNF2	QLAEQVVGAGEEWLTELDTDQLRNLLILDRSAVIDEEAE	1078
Nostoc_SNF2	QLAEQVVGAGEEWLTELDTDQLRNLLILDRSTVIDEEAD	1075
Nodsp_SNF2	QLAEQVVGAGEEWLTEMNTDQLRDLLILDRSAIIDEDEV	1087
Lyn_sp_SNF2	VLAEQVVGSGENWLTTELDTDQLRNLLIIDRNAVIDEEE	1061
Crowa_SNF2	KLAEQTVDAGEQWLTELDTDQLRNLLLLDRDTIIDEQ	1046
Ssynel_PCC6301_SNF2	ALAQQIVSGEDWLTELDTNQLRQLLILDRSAWVEEEEP	1019
Ssynel_PCC7942_SNF2	ALAQQIVSGEDWLTELDTNQLRQLLILDRSAWVEEEEP	1019
Theel_BP-1_SNF2	ALAEQVVGSGEHLTELNLDQLRQLLTLDKERLITL	999
Glovi_SNF2	ALAEQVVSAGENWLSDLNTDQLRQLLVLDSEIIDTETA	1008
Proma_CCOMP1375_SNF2	ELAENIVSGESWLGQLSLEKLSSELVALDSNPEF	1062
Proma_MIT\9211_SNF2	KLAENIIGAGEDWLGKLGINELRELVSLEKES	1067
Proma_MIT\9303_SNF2	RLAEDIIIGSGEDWLGGLGVSQRLRELVLEDSD	1099
Proma_MIT9313_SNF2	RLAEDIIIGSGEDWLGGLGVSQRLRELVLEDSD	1099
Syn_sp_CC9311_SNF2	RLAEDIIIGSGEDWLGGLGEMGQKELVSLLEDNQ	1064
Syn_sp_WH\7805_SNF2	RLAEDIVGSGEWLGGFDMGQKELVSLLEDNETRNP	1070
Syn_sp_RS9916_SNF2	RLAEDIIIGSGEDWLGGLDMGQKELVSLDDNGSLSA	1065
Syn_sp_CC9605_SNF2	RLAEDVIGSGEDWLGSLGGDQRLDLVSLDT	1065
Syn_sp_WH\8102_SNF2	RLAEDVIGSGEDWLGCLAGDQRLNLVALEDT	1063
Syn_sp_CC9902_SNF2	RMAEDVIGSGEDWLGSLGGDQRLNLVALEDT	1063
Syn_sp_WH\5701_SNF2	RLAEDIVGSGEWLGGDPPGQRLDLVALEE	1037
Myctu_SNF2	ALADLVVTDGEGWLTELSTRDLREVFFALSEGAVGE	1013
Mycbo_SNF2	ALADLVVTDGEGWLTELSTRDLREVFFALSEGAVGE	1013
Nocfa_IFM\10152_SNF2	RLADLAVDAGENWITELGTEELRELFTLGAEAVGE	956
Myxxa_DK_SNF2	QLAEKVVGAGEHWVTELDTTALRELFLSLSEGAVDDGDAEGEDDARVRAP	1080
Synth_IAM14863_SNF2	ALAGQVVISGESWLGQSLSTEELRALIALDREV	989
Metac_C2A_SNF2	QVAENVVGTGEGWLTELSNEELKDILALREEAVGE	1078
Metma_Go1_SNF2	QVAENVVGTGEGDWLTELSNDELKDILALREEAVGE	1089
Pelph_BU-1_SNF2	TVAGQVLGTGEQWLTELSNNDLRKLIMLGQEAAMGE	1007
Archaeon\RC-I_SNF2	ALSANILGTGEDWITELSTEQLRDMVMLRWDEVADDG	1042
Nos_sp_PCC7120_SNF2\II	KLSSAVVGSDESWLTELDNEAFKKLIALNKSTIME	1055
Bacce_ATCC10987_SNF2	SLNNAVIT-SDSWMTTELSTDELKELLGV	918
Methu_JF-1_SNF2	TLAKEVLAQSDSEYLTNLSTKELLEIVSLRDSLFRGEDA	1048

FIGURE 8 (continued)



40/96

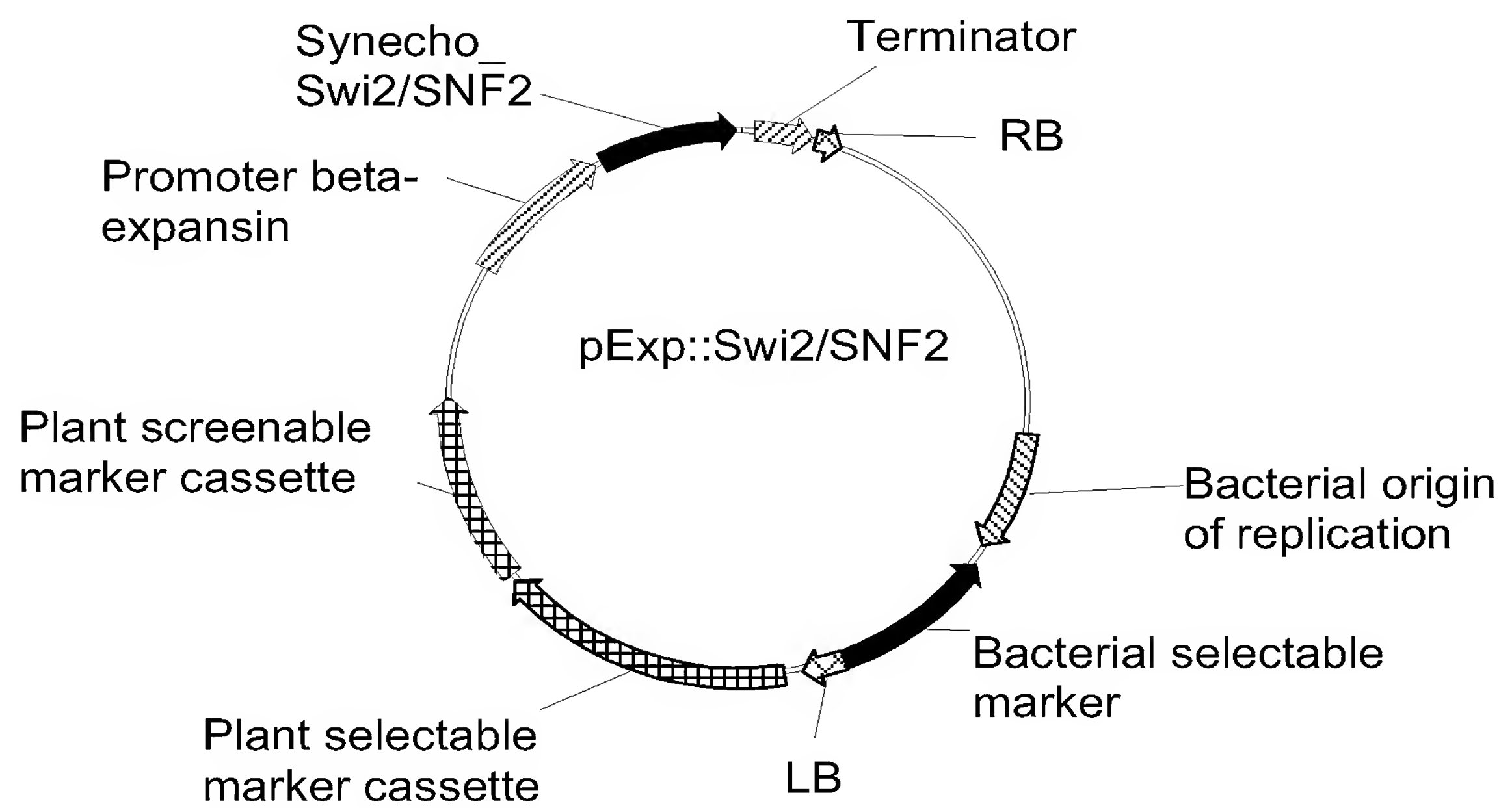


FIGURE 9

41/96

SEQ ID NO: 29, *Synechocystis* sp. PCC 6803 BA000022
Synecho_PCC6803_SNF2 nucleic acid sequence

TGTTTCGTTGCACAAATTGATGAGCAATGCTTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCT
TAAACAATGGCGACTATCCACGGTAATTGGCAACCCTCCCACGGGGGAAAACGGCGGCAGCAACTGTTT
CTTTGGGCGGATACCTGGGGTCATCCTTTGCCAGAAACCATTTGGCGATCGCCATCCCTTTGCGTTG
GATCTGCCGGATTTGCTACAGGCCTGGTTCGAATTTGCCCTGGCCTTCCCCAAGGCGGATGGGGTG
ACAGAGGCAGCCCTTACTCTGCATTTACCCAGCCATCGCCAGCAAAAAATTTCCCCTACCCTTTGTC
ACAGGGCAAGATCCGGTGGCCATGGATGCGAAATATCTCCACTGGCGATCGTGGCAGGTAACCGGG
GTAAATCTGACCCCAAGCCAAACGTTAACGTTGCTCCAATCTATTCCCCTGGGGGGGCCAAGCCTTA
GCTAACTTAGGATCAGAGTTTTACTTTTACGGTCAACTGCACCGCTGGTGTTTAGATTTGGTGCTA
CGGGGTAAATTTGTGCCGGGACTGGAGCAAAGGGGGGGAAGACGGTAATTACTATGCCCAATGGATT
CCTATCCTCGATAGCATCCAAGACCAAACCCATTTAGCCCAATTTAGCCAGAGAGTACCTGCCTGC
GCCCTGGCCAACCTGACTGACTCCCAGGAGCCCCAAATGTTGGTGGTGGATTTACTACAAAATTA
TTGCAAGCCCAAATTTGGTGCCGTCAGTCCCAGCCTAGCCAAACGTTAAAGAAGTCTGGTTGAATGAT
TGGCTCCGGGGGATTAACCCATGGGGGGGCAAACCTCCCTCGGCACAAGCAAAGCTCTACAACGATTA
GCCACATCCTTAGACCATTTGGTATTTACCAGTCCAGAATTATTTGGGCCAAAAAAATAACCAAGCT
TTAGCCCAACGGCAATGGCGGGGGGCTCTGCGGTTACAACCTCCAGCGGACGATGGGGGGGGAACC
TGGCAACTGGATTATGGTTTACAAGCCCTGGATGACGGGGAATTTTGGCTCCCGGCGGCTTCCCTC
TGGGCCATGGCCGGCGATCGCCTGGTGTGGCAGGGAAGGAGGGTTGACCAGGGGGGCGGAAAGTTTA
CTGCGGGGCTTAGGGGTAGCTGCCCAAATTTACGAACCCATTGCTGCAAGTTTGACGGAAAGGTGT
CCCACGGGCTGTGGGCTAGATGCCATCCAAGCCTACGAATTTATCCTGGCAATCGCCCATCAATTG
CGGGATCGGGGGTTAGGGGTAAATCCTCCCGCCGGGGTTAGAACGGGGCGGCACCGCCAAACGGTTA
GGGGTAAAAGTGGTGGGGGAAGTGCAACGGCAAAGGGGGCCAGCGGCTAACTCTGCAAAGTTTAATT
AATTACGACTTGCAACTAATGATGGGGAGCGGGGACAATGCCCGGTTATTGACGGCCAAGGACTTT
GAAGCGTTACTAGCCCAAAAATCTCCCCTGGTGGTGTGCTGGACGGAGAATGGATTACCCTGCAACCG
GCGGACGTGCGGGCGGCCAAGGTCATTTTACAGCAGCAACAATCTGCCCCGCCCTCACAGTGGAG
GATGCTCTGCGCCTCAGCATTGGTGTATTTACAAACCGTCTCTAAACTGCCGGTGACCCAGTTTGCT
GCTCGGGGCATATTACAGGAATTGATCGACACCCTCCGTAACCCGGAAGGAGTGAAAGCCATTGCT
GACCCACCGGGCTTTTACAGGGTACTTTACGGCCCTACCAAGCTCGGGGAGTGGGCTGGTTAGCTTTT
CTGGAACGGTGGGGGGCTGGGGGGCTGTTTGGCAGACGATATGGGTTTGGGAAAAACACCCCAGTTG
CTGGCTTTTCTGCTCCATTTAGCCGCGGAGGATATGTTAGTTAAGCCGGTGTGATTGTTTGTCTCT
ACGTCGGTGCTGAGCAATTGGGGTCATGAAATTAATAAGTTTTCGCCCCAACTTAAAACCCTATTG
CACCATGGCGATCGCCGGAAAAAAGGGCAACCGTTGGTTAAACAGGTCAAAGACCAGCAAATTGTC
CTCACCAGTTACGCTTTACTGCAACGGGATTTTAGTAGTTTGAATTTGGTGGACTGGCAGGGGATC
GTGCTGGACGAAGCCCAAAATATCAAAAATCCCCAAGCTAAACAGTCCCAGGCGGCCCGGCAATTG
CCAGCGGGTTTTTCGCATTGCCCTCACGGGGACTCCGGTGGAAAATCGCCTGACGGAATTGTGGTCA
ATTTTAGAATTTTTTAAATCCCGGTTTTCTGGGTAAATCAGAGCTTTTTTCCAACGGCGCTTTGCCAAT
CCCATCGAAAAATTTGGCGATCGCCAGTCGTTGTTAATTTTTCGCGGAATTTAGTGCGGGCCGTTTATT
TTGCGGCGGTTAAAAACCGACCAAACCATTTATTCAAGATTTACCAGAAAAACAAGAAATGACCGTC
TTCTGTGACCTTTCCCAAGAGCAAGCTGGTTTATATCAACAATTGGTGGAGGAATCCCTCCAGGCG
ATCGCCGACAGCGAAGGCATTCAAAGGCACGGTTTAGTTTTTAACCCTATTAACCAAACCTCAAACAG
GTTTGTAAACCATCCCGATCTATTGCTGAAAAAGCCCGCCATCACCCACGGGCACCAAGTCCGGCAAG
CTAATTCGTCTGGCGGAAATGCTGGAAGAAATCATCAGCGAAGGCGATCGGGTGTTAATTTTCACC
CAATTTGCCAGTTGGGGTCATTTACTCAAACCCTATCTGGAAAAATACTTTAACCAAGAGGTGCTC
TATCTCCACGGGGGCACTCCAGCAGAGCAACGGCAAGCTCTGGTGGAACGATTCCAACAGGACCCC
AACAGTCCCTATTTATTTATCCTTTCTCTCAAGGCTGGCGGCACAGGGTTGAACCTCACGAGGGCT
AACCATGTGTTCCATGTGGACCGGTGGTGGAAATCCGGCGGTGGAAAATCAGGCTACCGATCGTGCT
TTTCGCATTGGCCAAACTCGCAACGTCCAGGTGCACAAATTTGTCTGTACAGGCACCTTGGAAGAA

FIGURE 10

42/96

AAAATTAACGCCATGATGGCGGATAAACAACAATTGGCAGAACAAACCGTGGATGCCGGGGAAAAT
TGGCTCACCCGCCTAGACACCGATAAACTCCGTCAGTTGCTTACCCTCTCCGCCACCCCGGTGGAT
TACCAAGCCGAAGCGTCCGATTGAACCCAGCTTTCTTGTACAAAGTTGGCATGATAAGAAAGCATT
GCTTATCAATTTGTTGCAACGAACAGGTCCTATCAGTCAAAATAAAT

**SEQ ID NO: 30, *Synechocystis* sp. PCC 6803 BA000022
Synecho_PCC6803_SNF2 translated polypeptide**

MATIHGNWQPSHGENGGKLFLLWADTWGHPLPETIGDRHPFALDLPDLLQAWSNLPLAFPKADGVTE
AALTLLHLPSPHRQQKIPLPFVTGQDPVAMDAKYLHWRWSQVTGVNLTSPQTLTLLQSIPLGGQALAN
LGSEFYFYGQLHRWCLDLVLRGKFVPGLEQRGEDGNYYAQWIPILDSIQDQTHLAQFSQRPACAL
ANLTDSQEPQMLVVDLLQKLLQAQIGAVSPSLANVKEVWLNDWLRGLTHGGQTSLSGTSKALQRLAT
SLDHWYLPVQNYLGQKNNQALAQQRQWRGALRLQPPADDGGGTWQLDYGLQALDDGEFWLPAASLWA
MAGDRLVWQGRRVDQGAESLLRGLGVAAQIYEPIAASLTERCPTGCGLDAIQAYEFILAIHQLRD
RGLGVILPPGLERGGTAKRLGVKVVGVEVQRQRGQRLTLQSLINYDLQLMMGSGDNARLLTAKDFEA
LLAQKSPLVVDLGEWITLQPADVRAAKVILQQQQSAPPLTVEDALRLSIGDLQTVSKLPVTQFAAR
GILQELIDTLRNPEGVKAIADPPGFQGTLRPYQARGVGWLAFLERWGLGACLADDMGLGKTPQLLA
FLLHLAAEDMLVKPVLIVCPTSVLSNWGHEINKFAPQLKTLHHGDRRKKGQPLVKQVKDQQIVLT
SYALLQRDFSSSLKLVLDWQGIVLDEAQNIKNPQAKQSQAARQLPAGFRIALTGTPVENRLTELWSIL
EFLNPGFLGNQSFQRRFANPIEKFGDRQSLILRLNLRPFILRRLKTDQTI IQDLPEKQEMTVFC
DLSQEQAGLYQQQVEESLQAIADSEGIQRHGLVLTLLTKLKQVCNHPDLLLKKPAITHGHQSGKLI
RLAEMLEEI ISEGDRVLI FTQFASWGHLKPYLEKYFNQEVLYLHGGTPAEQRQALVERFQQDPNS
PYLFILSLKAGGTGLNLTRANHVHFHVDNRWNPVENQATDRAFRIGQTRNVQVHKFVCTGTLEEKI
NAMMADKQQQLAEQTVDAGENWLTRLDTDKLRQLLTLSATPVDYQAEASD

**SEQ ID NO: 31, *Anaebena variabilis* ATCC 29413 *Anava_SNF2* nucleic
acid sequence**

ATGGCAATTTTACACGGTAGTTGGATATTAAGTGAGCAGGATAGTTATTTATTTATTTGGGGGGAA
ACTTGGCGATCGCCACAAGTAAATTTTAGTTTTGAGGAAATAGCCCTCAATCCCTTGGCTCTGTCT
GCATCTGAATTAAGCGAGTGGTTGCAGTCTCAACATCAGGCGATCGCTCAGATTTTACCACAACAG
TTGGCAAAAAAACCTCCAAAGCAGCAAGTTCCCCAACAAACAAATTTACCAATTCCTCGCAAATA
ATTGTTCTGCCAACGGAAATTTCTCAACCTCGTAAGAAAGAAACAATTTTCATTTCTCCTGTGCAT
TCTGCCGCTTTAGAATCTGATGCAGACTCTGAAGTTTATTTACAACCTTGGCGTGTAGAAGGTTTT
TGTCTTCCTCCTAGTGCAGCAGTTAAATTTCTAACTTCTTTACCTTTAAATATCACTAGCACAGAG
AATGCTTTTTTTAGGTGGAGATTTACGTTTTTGGTCACAAATTGCCCGTTGGAGTTTAGATTTAATT
TCTAGGTCTAAGTTTCTCCCAATTATCCAACGACAACCTAATAATTCTGTAAGTGCCAAATGGCAA
GTACTGTTAGATAGTGCTGTAGATGGAACCTCGTTTAGAAAAGTTTCGCCGCGAAGATGCCTTTGGTT
TGTCGGACTTATCAGAGATTAGGGAACGAGGAATTATCTCCATCTCCTATATATATAGATTTTCCT
AGTCAGCCGCAGGAATTAATATTGGGTTTTCTCAATAGTGCAATAGATACGCAATTACGGGAAATG
GTGGGGAATCAGCCTGTGGTGGAACTCGCTTGATGGCATCTTTACCGTCGGCGGTACGACAGTGG
CTGCAAGGGTTAAGTGGTGCATCTAATTCAGTTGATGCAGATGCAGTTGGTTTGGAAAGGCTGGAA
GCAGCGCTCAAGGCTTGGACGATGCCGCTACAATATCAACTAGCAAGTAAAAATCAATTTTCGCACC
TGTTTTGAATTACGTTCTCCAGAACCAGGAGAACTGAATGGACACTAGCCTATTTCTGCAAGCA
GCCGATAATCCAGAATTTCTAGTAGATGCGGGCACTATTTGGCAACATCCTGTTGAACAGCTAATT
TATCAACAGCGATCGATTCAAGAACCCAGGAAACATTTTTTACGAGGTTTGGGGTTAGCTTCTCGA
TTGTATCCGGTCATTGCCCCCACTTTAGATACAGAATCACCGCAATTTTGTTCATCTCAACCCCATG
CAGGCTTATGAATTTATCAAGGCTGTGGCTTGGCGATTTGAAGATAGCGGTTTAGGGGTGATTTTA
CCTCCTAGTTTGGCGAACCAGGGAAGGCTGGGCAAACCGCTTGGGATTGAAAATCTCCGCCGAAACC
CCAAAGAAAAAGCCAGGACGCTTGGGATTGCAGAGTTTGCTTAATTTTCAATGGCACTTAGCAATT

FIGURE 10 (continued)

43/96

GGTGGGCAAAC TATTTCTAAAGGGGAATTTGACAGACTAGTAGCTTTAAAAAGCCCATTGGTAGAA
ATAAATGGCGAATGGGTGGAGTTGCGTCCCCAAGATATCAAGACAGCCGAAGCCTTTTTTTGCTGCA
CGTAAAGACCAAATGGCCTTATCTTTAGAAGATGCTTTACGTCTGAGTAGTGGGGATACTCAAGTA
ATTGAGAAATTACCAGTAGTCAGCTTTGAAGCCTCTGGCGCATTACAAGAATTAATTGGGGCGCTG
ACAAATAATCAAGCAGTTGCACCATTACCTACGCCAAAGAACTTCCAAGGAAAGTTGCGTCCTTAT
CAAGAAAGGGGTGCGGCTTGGTTGGCATTCTCTCGAACGCTGGGGTTTAGGTGCTTGTCTCGCCGAC
GACATGGGACTGGGAAAAACGATACAGTTTATTGCTTTCTCTCCATCTTAAAGAACAGGATGTA
TTAGAAAAACCAACTTTACTAGTGTGTCCTACTTCTGTTTTAGGTAACTGGGAACGAGAAGTGAAA
AAATTTGCACCTACACTTAAAGTTCTCCAATATCATGGTGATAAACGTCCTAAAGGTAAAGCTTTT
CCAGAAGCAGTAAAAAATCATGATTTAGTTATCACCAGTTACTCACTAATTCATAGAGACATCAAA
TCATTGCAGGGTCTTTCTTGGCAGATAATTGTTTTAGATGAAGCCCAGAATGTGAAGAATGCGGAA
GCCAAACAATCACAAGCAGTCCGACAATTAGACACAACCTTTTCGCATTGCTTTAACGGGGACACCA
GTCGAAAATAGACTACAGGAACCTTTGGTCAATTTTAGATTTCTCAACCCTGGTTATTTAGGTAAT
AAGCAATTCTTCCAAGACGCTTTGCCATGCCAATTGAAAAGTATGGTGATGCAGCATCTTTAAAT
CAATTGCGTGCCTTAGTACAACCATTTATTCTGCGTCGCCTGAAAACAGACCGTGATATTATTCAA
GACTTGCCAGATAAGCAAGAAATGACAGTATTTTGCGGTTTGACTGGAGAACAAGCTGCACTTTAT
CAAAAAGTGGTAGAAACATCTTTAGCAGAAATTGAATCGGCCGAAGGATTGCAACGCCGAGGGATG
ATTTTAGCTTTATTAATTAACTCAAACAAATCTGCAATCATCCAGCCCAATATCTGAAAACAAAT
ACCTTAGAACAAATACAGTTCAGGAAAACCTGCAACGATTAGAAGAAATGTTAGAAGAGGTGTTAGCG
GAGAGTAATACTTATGGTGTTGCTGGTGCGGGACGTGCTTTAATCTTCACCCAGTTTGCAGAATGG
GGTAAGTTACTCAAACCACATTTAGAAAAACAACCTAGGGCGGGGAAGTATTTTTCTTATATGGTAGT
ACCAGTAAAAAGCAACGTGAAGAAATGATTGACCGTTTTCAACACGACCCCTCAGGGGGCCACCAATT
ATGATTCTCTCTCTCAAAGCAGGTGGTGTAGGGTTGAACTTAACCAGAGCAAATCATGTATTTTAC
TTTGATAGATGGTGGAATCCAGCCGTAGAGAACCAAGCCACAGACCGCGTATTTTCGTATTGGTCAA
ACCCGCAATGTACAGGTGCATAAATTTGTTTGCAATGGTACCTTAGAAGAAAAAATCCACGACATG
ATTGAAAGTAAAAACAACCTAGCGGAACAGGTGTTGGTGCAGGCGAAGAGTGGTTAACTGAATTA
GATACAGATCAACTCCGCAACTTACTGATACTTGATCGTAGTGAGTAATTGATGAAGAAGCAGAG
TAA

SEQ ID NO: 32, *Anaebena variabilis* ATCC 29413 Anava_SNF2 translated polypeptide

MAILHGSWILSEQDSYLFIWGETWRSPQVNFSFEEIALNPLALSASELSEWLQSQHQAIQAQILPQQ
LAKKTSKAASSPTTNLPIHSQIIVLPTEISQPRKKETIFISPVHSAALESADSEVYLQPWVEGF
CLPPSAAVKFLTSLPLNITSTENAFLLGGDLRFWSQIARWSLDLISRSKFLPIIQRQPNNSVSAKWQ
VLLDSAVDGRLEKFAAKMPLVCRTYQRLGNEELSPSPIYIDFPSQPQELILGFLNSAIDTQLREM
VGNQPVVETRLMASLPSAVRQWLQGLSGASNSVDADAVGLERLEAALKAWTMPLQYQLASKNQFRT
CFELRSPEPGETEWTLAYFLQAADNPEFLVDAGTIWQHPVEQLIYQQRSIQEPQETFLRGLGLASR
LYPVIAPTLDTESPQFCHLNPMQAYEFIKAVAWRFEDSGLGVILPPSLANREGWANRLGLKISAET
PKKKPGRLGLQSLNLFQWHLAIGGQTI SKGEFDR LVALKSPLVEINGEWVELRPQDIKTAEAFFAA
RKDQMALSLEDALRLSSGDTQVIEKLPVVSFEASGALQELIGALTNNQAVAPLPTPKNFQGKLRPY
QERGAAWLAFLERWGLGACLADDMGLGKTIQFIAFLLHLKEQDVLEKPTLLVCPTSVLGNWEREVK
KFAPTLKVLQYHGDKRPKGKAFPEAVKNHDLVITSYSLIHRDIKSLQGLSWQIIVLDEAQNVKNAE
AKQSQAVRQLDITTFRIALTGTPVENRLQELWSILDFLNPGLGNKQFFQRRFAMPIEKYGDAASLN
QLRALVQPFILRRLKTDRIIQDLDPKQEMTVFCGLTGEQAALYQKVVETSLAEIESAEGQLQRRGM
ILALLIKLKQICNHPAQYLKTNLTLEQYSSGKLQRLEEMLEEVLAESENTRYGVAGAGRALIFTQFAEW
GKLLKPHLEKQLGREVFFLYGSTSKKQREEMIDRFQHDPPQGPIMILSLKAGGVGLNLTRANHVVFH
FDRWWNPAVENQATDRVFRIGQTRNVQVHKFVCNGTLEEKIHDMIESKKQLAEQVVGAGEEWLTEL
DTDQLRNLLILDRSAVIDEEAE

FIGURE 10 (continued)

44/96

SEQ ID NO: 33, uncultured methanogenic archaeon RC-I Archaeon_RC-I_SNF2 nucleic acid sequence

ATGATTACACTTCACGGAACCTGGACTACTGTTCGATCCCCTGAATGGCACATTTTTCTCTGGGGA
GAGAGTGATCCGGCCACGCAGCATAAAAGAAGAGGCAGGCCTCGGAAAAGTGCAGGGGAGAAACAG
CACCCGTTTTCACGCCGGCATCAAAGAGCTGGAAGCTGGAGCGGGGGCTATCAATTCATCGTGTATA
AGACATATAGCAGATGCGGGAGCACGGGCGGAGCAGGTTTTAATTTTGCCGTCAGCTACGGACAGG
CCCCTGAGATCTGCGAGCCCTTCAGCACTGGAGTCAGGTGAAGAAACCAACCCTGACAGCAGTTTA
CAATTTCTTCCGTGGACGGTGACCGGCATCAACATTAAGCCCGGGAATGCTCTGGTACTTCTATCC
TCTATAGCCGAATCACAAAAGCGGATCGGAGATATGGCGATAGGCCCAGACCTGCTTTACTGGAGT
AAGGTAGCCAAGTTTACGCTTAAGCTCCTGATAAGCCAGCAGTTCAGGCCGGAGGTTGTCGAAGTA
ATGAGCGGAAAAGCATATAGCCGTTGGAGATTTGCGCTCACCGATGAAACTGACCGGAAACACTAT
GCCTCGCTCGAAAACCTCATGCCGCTGGCATGTATTGCGGTTTCAGGAAAGGCTGGCATTTATAAT
CGAAAAGAAGCCTTAGATTTGTTCATTAATAACCGCCCTTGACACATTTATCCGGGACCAGATTGCC
CTGCCCCGCTGACAGCAGGATGACGAACCTGCTATCGCAAGCATGGCTAGATTGCTCGGCACCGGA
GAGAGTATCCGCCTGTCGGCTCCTGAGATGAAGAACTCAAAGATTCGGCAGGCCGCTGGACATCC
CGCATGAAAACAGAGAGCAAACAAGCTTTAAAGACCTGCTTCATCCTGGAGCCGCCAGCCCCGGAT
ACAGAGTATCCTGAAGCGCCGTGGAACCTACGGTACTGCTTGCAGGCATCCGATGACCCCAGTCTG
GTAATTCCGGCTGAGACTGTGTGGAAAGAGTTGAAGAAGACGCTGAAGTACCTGAATAAGAGATAC
GATAACCCTCAGGAGCAATTGTTACAGGATCTCGGAAAAGCGATGCAGATGTTTCCCGAAATCGAG
CCCAGCCTCAACACGTCAAACCTCTGTCCGCAACGCTGAGCACCAGTGAAGCCTACAAGTTCCTG
ACAGAAGCGGCGCCTCTGCTGCAGGACAGCGGGTATAGCATTATCCTACCGGAATGGTGGCGCAAC
AGCACTGGCAGGCTCAAGCTCGGCGCCAGGCTTCGCTTCAAGCCGAAAGCCGAAGGTAAAGCGGGT
AAAAGCCAGTTCACCATGGATAACCCTCGTCAGCTACGACTGGCGCCTGGCGCTGGGCGATCAGGAG
ATCACCGAAACAGAGTTCAGGAAGCTGGCAGCCCTGAAAGAGCCGCTTCTGCAGATAGGCGGGAAA
TGGTTTTCGCTGAAAAAGGAAGACATAGACAGCATCATGAAAGCATTTCAGGGCGAAGAAGACTGGA
GAGATGGCTTTATCGGAGGCACTGCGCCTCAACGGCGGGCTGGAAGACTTCAACGGCATCCCCGTC
AGCGGCATGAAATCGTCAGGATGGCTGGCAGAACTTTTCGACAGGCTGGCAGCCGGCGAAAAAATA
ACGAGCCTTGCCCCGCGGACGGTTTCAACGGGGAGCTTAGAGATTACCAGGTTAAAGGCTACTCC
TGGCTGGCCTTCATGAAAAAGTATGGCCTGGGCTCCATTCTGGCTGACGACATGGGCCTGGGTAAAG
ACGATACAGCTGCTGGCGTTGCTCCTGAAAGAGAAGGAAAGAGGCACTAAAGGCCCTACTCTGTTG
ATCTGCCCCACCTCGATTCTCGGAAACTGGCAGCGGGAGGCGAAGAAATTTGCCCCGGCCCTGAAA
GTCCACATACACCATGGGGCAGGAAGGGCTGATAAAGAGCAGTTCGGAAAAATCGTCAAGGCTCAC
GACCTGATCCTGAGCACTTACGCTCACGCCTACCGGGACGAGGAACTGCTTAAAGAGGTGAACTGG
AAGCTGGTAGTGCTCGACGAGGCTCAGAATATCAAGAATCATCATACCCGGCAGGCCAGAGCTATC
CGGGCTCTTAAGGCCGATCACCGAATAGCCATGACGGGAACGCCGATAGAGAACAGACTCTCGGAG
CTGTGGTCGATCGTGGACTTCCTGAACCCCGGCTACCTGGGCAAGGCGGAGACATTCAGGAAACAA
TTCGCCATACCTATCGAGAGATACGATGACGCTGCCCGGTTCGGAAAAATGAAGCAGGCCATCAAG
CCCCTGGTGCTGCGCAGAGTGAAGACGGATCCGGCCATCATCAAAGACCTGCCGGACAAGATCGAG
ATCAAGGAGCCCTGCAACCTCACCAAAGAAGAGGCCACGCTCTACGAGGCCATCGTAGAGAACATG
CTGAAAAGTATAGATAAGGCCACGGCAATGCAGAGACGGGGAATCGTCTTAGCGTCCCTGATGAAG
CTCAAACAGGTCTGCGATCACCCGTCGCTGTACATCAAACGGGCGCTGTGACCGACGATAAGACG
CTGATCAGGTCTGGCAAGCTGAAGCGCCTCACGGAGCTGCTCGAAGAAGCGCTGGCCGAAGGCGAC
AGCGTGCTGATCTTCACCCAGTTCGTGGAAATGGGGGAGATGCTGAAAGCCTACCTGCAGAGCACG
TTCGACGAAGAAGCCCTCTTTTTTGCACGGCGGAGTACCGCAGAAGGCCAGAGACAAGATGGTCCTC
CGTTTTCGGGGAAAAGGACGGGCCACGGATCTTTATCGTCTCGCTGAAAGCCGGCGGCGTCCGGCCTC
AACCTGACGAAGGCAAGCCACGTGTTCCACTTCGATCGCTGGTGGAACCCGGCGGTTCGAGAACCAG
GCGACAGATCGAGCTTACAGGATAGGCCAGAGCAAAAATGTACTGGTCCATAAATTCGTCTGCGCC
GGCACGCTGGAAGAAAAGATCGACGAGCTGATCGAGAGCAAAAAGGCGCTGTCCGGCGAACATCCTC
GGCACGGGAGAAGACTGGATCACGGAGTTGTGACCGAACAGCTGAGGGACATGGTCATGCTGAGA
TGGGACGAGGTAGCCGATGATGGCTAA

FIGURE 10 (continued)

45/96

SEQ ID NO: 34, uncultured methanogenic archaeon RC-I Archaeon_RC-I_SNF2 translated polypeptide

MITLHGTWTTVDPLNGTFFLWGESDPATQHKRRGRPRKSAGEKQHPFHAGIKELEAGAGAINSSCI
RHIADAGARAEQVLILPSATDRPLRSASPSALESGEETNPDSLSLQFLPWTVTGINIKPGNALVLLS
SIAESQKRIGDMAIGPDLLYWSKVAKFTLKLLISQQFRPEVVEVMSGKAYSRWRFALTDDETDRKHY
ASLENSMPLACIAVSGKAGIYNRKEALDLFINTALDTFIRDQIALPADSRMTNLLSQAWLDSLGTG
ESIRLSAPEMKKLKDSAGRWTSMKTESKQALKTCFILEPPAPDTEYPEAPWNLRYCLQASDDPSL
VIPAETVWKELKKTLKYL NKRYDNPQEQLLQDLGKAMQMFPEIEPSLNTSKPLSATLSTSEAYKFL
TEAAPLLQDSGYSIILPEWWRNSTGRLKLGARLRFKPKAEGKAGKSQFTMDTLVSYDWRLALGDQE
ITETEFKRLAALKEPLLQIGGKWFALKKEDIDSIMKAFAKKTGEMALSEALRLNGGLEDFNGIPV
SGMKSSGWLAEFLDRLAAGEKITS LAPDGFNGELRDYQVKGYSWLAFMKKYGLGSILADDMGLGK
TIQLLALLLKEKERGTGKPTLLICPT SILGNWQREAKKFAPALKVHIHHGAGRADKEQFGKIVKAH
DLILSTYAHAYRDEELLKEVNWKLVLDEAQN IKNHHTRQARAIRALKADHRIAMTGTPIENRLSE
LWSIVDFLNPGLGKAETFRKQFAIPIERYDDAARSEKLKQAIKPLVLRVKTDPAI IKDLDPKIE
IKEPCNLTK EQATLYEAIVENMLKSIDKATAMQRRGIVLASLMKLKQVCDHPSLYIKTGAVTDDKT
LIRSGKLRLELLEEALAE GDSVLI FTQFVEMGEMLKAYLQSTFDEEALFLHGGVPQKARDKMVL
RFGEKDGPRIFIVSLKAGGVGLNLTKASHVFHFDRWWNPAVENQATDRAYRIGQSKNVLVHKFVCA
GTLEEKIDELIESKKALSANILGTGEDWITELSTEQLRDMVMLRWDEVADDG

SEQ ID NO: 35, Bacillus cereus ATCC 10987 Bacce_ATCC10987_SNF2 nucleic acid sequence

ATGATCAATCAAACCTGAAGTAACAATTAGGCTCCAGCACGTTAGTCACGGTTGGTTCCTTTGGGGA
GAAGATGATAGCGGTACTCCATTATCCGTAACAAGTTGGAAACGAAATGCATTTACATGGCACTCC
ACTTCCTTCTACGGCACGTTTCTAAAAGAAGCAAGCTTTGAAGGAAGACAAGGTGTTATGCTAACA
AACGCACAAGCATTGTGAATACATCGCGAATAAACCGATGAACTCCTTTGCCCGTATTCAAATGAAC
GGCCCTATTACAGCACTTACGGAAGATGCGAACGAATTGTGGGATGCCTTCACAAGCGGTAGCTTC
GTACCTGATATGGAGCGTTGGCCTAAACAACCATCTTGGAAAGTTCAAATACTCCAATCGAAGAT
GAAACATTGGCATCTCTTTTCTCGGCTGCAGTAAATGAAAGCATATTACAAGATAACCGTTCAAAT
GACGGATGGGAAGATGCAAAGAGACTTTATGAACATTACGACTTTACGAAAAGACAATTAGACGCA
GCACTACATGAAGAAGATTGGCTTCGAAAAATTGGTTACATTGAAGATGACCTTCCCTTTACAATC
GGACTACGACTACAAGAGCCGCAAGAAGAATTTGAAATGTGGAAGCTTGAAACAATTGTTACGCCA
AAGCGCGGGGCACATCGCATATATGTATATGAGAGTATCGATTCTTTACCAAACGATGGCACGAT
TATGAAGAACGTATTCTGGAAACACAAGAAAGCTTCAGTAAGCTCGTACCGTGGCTAAAAGATGGT
GATACATTCCGAAGTGAACCTTTGAAACAGAAGCGTGGAACCTTCTTAACAGAAGCAAGTAACGAA
TTACTCGCCGCAGGTATTACAATCTTATTACCATCGTGGTGGCAAAATTTAAAAGCGACAAAACCA
AAATTACGTGTGCAACTGAAGCAAAATGCTACACAAACGCAATCTTTCTTCGGCATGAATACACTC
GTTAATTTTGA CTGGCGCATTTCAACGAACGGCATTGATTTATCAGAAAGCGAATTTTTTGA ACTC
GTTGAACAAAACAAGCGGTATTCAATATAAATGGTCAATGGATGCGACTAGATCCAGCCTTTATT
GAAGAAGTACGAAAGCTCATGAATCGTGCTGATAAGTATGGACTTGAAATGAAAGATGTCCTGCAG
CAACATTTTATCAAACACGGCTGAAACAGAAATTGTAGAAGAGGATAGTCCGTTTACAGATATTGAA
ATTGAACTAGATGGATATTATGAAGACTTATTCCAAAACTATTGCACATTGGAGATATTCCGAAA
GTAGATGTCCCTTCATCACTAAACGCCACACTCCGTCCGTATCAACAACATGGCATTGAGTGGTTA
TTATATTTAAGAAAGCTTGGATTCTGGCGCATTTGTTAGCTGACGACATGGGACTTGGAAGAGTATT
CAAACGATCACTTACTTACTATATATAAAAGAAAACAATCTCCAAACAGGTCCTGCTTTAATCGTG
GCTCCGACATCTGTTCTTGGAAATTGGCAAAAAGAATTTGAGCGTTTCGCACCGAATTTACGTGTT
CAGTTACATTATGGAAGTAACCGAGCTAAAGGGGAACCCTTTAAAGATTTCCCTTCAATCAGCAGAT
GTTGTATTAACATCTTATGCATTAGCTCAGCTTGATGAGGAAGAACTTAGTACGTTATGCTGGGAT
GCTGTTATTTTGGATGAAGCACAAAATATTAAAAACCCACATACGAAACAGTCTAAAGCAGTACGA

FIGURE 10 (continued)

46/96

AACTTACAAGCAAATCACAAAATCGCATTAACTGGGACACCGATGGAAAACCGCCTTGCCGAGCTT
TGGTCTATTTTCGACTTCATTAATCATGGATATCTTGGCAGCTTAGGACAATTCCAGCGCCGCTTC
GTCTCACCAATTGAAAAGGACCGTGACGAAGGAAAAATCCAACAAGTTCAACGTTTTATCTCACCG
TTTTTACTGCGTCGTACGAAGAAAGATCAAACAGTCGCATTAACTTACCAGATAAACAAGAACAG
AAAGCTTACTGTCCACTAACTGGTGAACAAGCTTCCTTATATGAACAACCTTGTTCAAGATACGTTG
CAAATGTAGAAGGATTAAGCGGAATTGAACGACGCGGATTTTATATTACTCATGCTGAACAACTT
AAACAAATTTGTAATCATCCCGCTCTTTATTTAAAAGAAACAGAACCGAAAGACATCATCGAGCGT
TCCATGAAAACGAGCACGCTCATGGAACCTATTGAAAATATAAAAGATCAAATGAAAGTTGCTTA
ATCTTCACGCAATACATCGGTATGGGGAACATGCTAAAAGATGTGTTAGAAGAACATTTTCGGTCAG
CGCGTCCTCTTCTTAAACGGTAGTGTACCGAAGAAAGAACGTGACAAAATGATCGAACAGTTCCAA
AACGGAACGTATGACATCTTCATTTTATCGTTAAAAGCAGGTGGTACAGGATTAACTTAACAGCT
GCCAACCATGTCATTCACTACGATCGTTGGTGAATCCAGCGGTAGAAAACCAAGCAACAGACCGT
GCATATCGCATTGGTCAAAAGCGCTTCGTTACGTTTATAAACTGATTACAACGGGGACACTTGAA
GAGAAAATCGATGAAATGTTAGAAAGAAAACAATCATTAAACAACGCCGTCATTACAAGCGATAGT
TGGATGACAGAACTATCTACAGATGAACTAAAAGAATTACTTGGTGTATAA

**SEQ ID NO: 36, *Bacillus cereus* ATCC 10987 Bacce_ATCC10987_SNF2
translated polypeptide**

MINQTEVTIRLQHVSHGWFLWGEDDSGTPLSVTSWKRNAFTWHSTS FYGTFLKEASFEGRQGVMLT
NAQAFEYIANKPMNSFARIQMNGPITALTEDANELWDAFTSGSFVPMERWPKQPSWKVQNTPIED
ETLASLFSAAVNESILQDNRSNDGWEDAKRLYEHYDFTKRQLDAALHEEDWLRKIGYIEDDLPFTI
GLRLQEPQEEFEMWKLETIVTPKRG AHRIYVYESIDSLPKRWHDYEERILETQESFSKLPWLKDG
DTRSELFETEAWNFLTEASNELLAAGITILLPSWWQNLKATKPKLRVQLKQ NATQTQSFFGMNTL
VNFDWRISTNGIDLSESEFFELVEQNKRLFNINGQWMRLDPAFIEEVRKLMNRADKYGLEMKDVLQ
QHLSNTAETEIVEEDSPFTDIEIELDGYIEDLFQKLLHIGDIPKVDVPSSLNATLRPYQQHGIEWL
LYLRKLGF GALLADDMGLGKSIQTITYLLYIKENNLQTGPALIVAPTSVLGNWQKEFERFAPNLRV
QLHYGSNRAKGEPFKDFLQSADVVLTSYALAQ LDEEELSTLCWDAVILDEAQNIKNPHTKQSKAVR
NLQANH KIALTGTPMENRLAELWSIFDFINHG YLGSLGQFQRRFVSPIEKDRDEGKIQQVQRFISP
FLLRRTKKDQTVALNLPDKQE QKAYCPLTGEQASLYEQLVQDTLQNVEGLSGIERRGFILLMLNKL
KQICNHPALYLKETEPKDI IERSMKTSTLMELIENIKDQNESCLIFTQYIGMGNMLKD VLEE HFGQ
RVLFLNGSVPKKERDKMIEQFQNGTYDIFILSLKAGGTGLNLTAANHVIHYDRWWNPAVENQATDR
AYRIGQKR FVHVHKLITTTGTLEEKIDEMLERKQSLNNAVITSDSWMTLSTDELKELLGV

**SEQ ID NO: 37, *Crocospaera watsonii* WH 8501 ctg336 Crowa_SNF2
nucleic acid sequence**

ATGACAATATTACATGGAACCTTGATTGAAAATACCTCTGAAAAACATTTTTTTTATTTGGGGGGAA
ACTTGGCGTTCTTTATCCTCTGATATTTCTCAGATGATTCTATTTTAATGTATCCATTTTCTGTA
GATAAACAGGGAATTATTGAACAATTAACTCGAATAAGATTAAAGATTGAAAAAACAAAAATATT
GAATCTGTTTCTCAAATATTTTATTTGCCTAGTAAATTTATTGCTAAATCGAAGCAAAGTATCCCT
TTACTATCAACAGAATTAAAAGATAAAGATTTTGAACAAGGGGATATTCAGTTAATTGCTTGGAAA
ATCGAAGGGATAAAATTAATGTTGATGATACAATTAATATTTTAAGTCAGTTACCGTTGGGATTA
ACCAATAATGACGAAAATTACATAGGCGATAATTTAAAATTTTGGACACATATTTATCGTTGGAGT
CTAGATTTATTA ACTAGAGGTAAATATTTACCGCAAATGGAAGAACAAGATAATAACTGTTATGGA
CAATGGGAACCTTTACTAGATAGTTTAGTTGATCAGCAACGGTTCTCTAAATTTATACAAACTATG
CCAAATAGTTCTCTTGCTTATCATAATTTAATGGAGGGTGAATTATCCTCTTCTTTACTCAAACAA
ACTACTATTCTTGATTTTTTATCTACTATCATTAATCAACAAGTACGTCAATTTATTGATGTTGCT
ATTACCCCTAGTTCATTTATCCAAAAGTGGTTATACTCTTTAACACAAGACTTATCTAAATTTGAA
GCATCAGAAGTTGAAAGAAAGGGATTAAAGAATGCTATTAATAATTGGAAATCTTCTTTAAGTGAA

FIGURE 10 (continued)

47/96

TATATTATAAAGTCTGATAATCAACCATTAGGAATTAACCAGTTTCGTGTTTGTGTTTTAAACTAGAA
AATCCAGCTAAAAGTGGTAAGAAATTAGAACAAAGTAATTGGCAGTTACACTACTATCTCCAAGCT
TTAGATGATCCTAATTTTCTGATCTCTGCCAAGGTTATTTGGGAAAATCCTGTTACTAGATTAATC
TGCAATAATAGAACAAATTAATCATCCTCAAGAAACCTTGCTAAAAGGACTAGGTTTAGCTTCACGT
CTATATTATCTAATTGAAGAAAGTTTACAAGACAATAAGCCTAGTTTTTCTGAGTTAGATCCCATA
CAAGTCTATGAATTTTTTACGTTCAATTGCTAATATTCTTAAAGATAATGGCTTAGGGGTTATCTTA
CCAGCTAGTCTAGAGCAAGGAGTCGAAGAAAAACGCTTAGGAATTAGTCTAACC GCAGAAAGTTAAG
TCGAAAAAAGGACAAAGACTTAGCTTACAAAGTTTGTTAAGTTATAAGCTAAATTTAGCAATTGGT
GATAAAACAATATCGAAAAAAGACTTTGAAAAACTATTAGCGCAAAAGTCACCTTTAGTTGAAGTA
AAAGGAGAATGGATAGCATTACAACCTGCTGATGTCAAGGCCGCACAACAAATTTTAAATAAGTCC
TATGATCCCCCTAGAACTTTCTGTAGAAGATGCTTTACGCTTCAGCACAGGAGATATTTCAACTGTT
GCCAAACTGCCGATTACTAACTTTGAAGCAAAAGGGGAATTAGCCAATCTAATTAATGCTATAAAT
AATAATGAATCAATCCCTATGATCGAAAATCCCAGAGGATTTAAAGGTCAATTACGTCCCTATCAA
CAGCGAGGAGTCGGTTGGTTATCGTTCTTAGAAAAATGGGGTTTAGGGGCTTGTCTTGCCGATGAT
ATGGGATTAGGAAAAACACCACAATTAATTGGGTTTCTCTTACATTTAAGAAGCGAAGGAATGTTA
GATCAACCTACCTTAGTTATTTGTCCTACATCTGTTTTAAATAACTGGGAAAGAGAAGTTCAAAAA
TTTGCCCCAACCCCTTTCTACTTTGATTTCATCATGGAGATAAACGTAGTAAAGGGAAAGCTTTTGT
AAAGCAGTTAGTAAAAAAAATGTTATCATTACTAGCTATTCTTTAATTTATCGAGATATTAAAAGC
TTTGAACAGGTAGAATGGCAAGGTATTGTCTTAGATGAAGCACAAAATATAAAAAATCCCCAGGCA
AAACAATCCCCAAGCAGTGCGTCAAATTTCCACACAGTTTCGTATTGCTTTAACAGGAACTCCTGTA
GAAAATCGCCTAACAGAATTATGGTCAATTCTTGACTTTCTTAACCCAGGATTTTTTAGGGACACAG
CAGTTTTTCCGTCGTCGTTTTGCCACTCCTATCGAAAAATATGGGGATAAAGAATCACTGCAAATT
ATGCGTTCTTTGGTACGTCCTTTCATTCTCAGACGATTGAAAACAGATAAACTATTATTCAAGAT
TTACCCGAAAAACAAGAAATGACCATTTTTTTGTGGGTTATCCTCAGAACAAGGAAAACCTTTATCAA
CAATTAGTAGATAATTCTCTGGTAGCAATAGAAGAGAAAACAGGAATTGAACGCAAAGGCTTAATT
TTAAGCTTACTGCTAAAACTCAAACAAATTTGTAACCATCCTGCTCATTTTCTCAAGCAAAAGAGC
TTAAAAACAGCAGAACAATCTGGTAAATTATTAAGACTAGAAGAAATGCTAGAAGAATTAATCGAA
GAAGGAGATCATGCTTTAATCTTTACCCAATTTTCTGAATGGGGTAAACTGCTGCAACCTTATTTA
CAGAAAAAATTTCAAGACGTTCTCTTTTTGTATGGTGCTACTCGCAGAGTTCAAAGACAAGAA
ATGATCGATCGCTTTCAACAGGATCCCAACGGACCCAGAATTTTTATTCTCTCCTTAAAAGCAGGG
GGAACCGGATTAAATTTAACCCGCGCTAACCATGTATTTTCATATTGATCGTTGGTGGAACCCAGCA
GTAGAAAATCAAGCAACCGATCGCGCGTTTCGTTTAGGACAAAAACGCAATGTTCAAGTACATAAA
TTTGTCTGTACAGGAACCCTAGAAGAAAAAATTAACGAAATGTTAGAAAGTAAACAAAAAATTAGCC
GAACAAACCGTTGACGCAGGGGAACAATGGTTGACAGAATTAGATACAGATCAACTGCGTAACCTC
TTATTATTGGATCGAGATAACCATTATTGACGAACAATAA

**SEQ ID NO: 38, *Crocospaera watsonii* WH 8501 ctg336 Crowa_SNF2
translated polypeptide**

MTILHGTWIENTSEKHFFIWGETWRSLSSDISSDSILMYPFSVDKQGIIEQLNSNKKIKIEKNKNI
ESVSQIFYLPSKFIAKSKQSIPLLSTELKDKDFEQGDIQLIAWKIEGIKLNVDDETINILSQLPLGL
TNNDENYIGDNLKFWTHIYRWSLDLLTRGKYLPMEEQDNNCYGQWEPLLDLSDVQQRFQSKFIQTM
PNSSLAYHNLMEGELSSSLLKQTTILDFLSTIINQQVRQFIDVAITPSSFIQKWLYSLTQDLKFE
ASEVERKGLKNAINNWKSSLSEYIIKSDNQPLGINQFRVCFKLENPAKSGKKLEQSNWQLHYLQA
LDDPNFLISAKVIWENPVTRLICNNRTINHPQETLLKGLGLASRLYYLIEESLQDNKPSFSELDPI
QVYEFLRSIANILKDNGLGVILPASLEQGVEEKRLGISLTAEVKSKKGQRLSLQSLLSYKLNLAIG
DKTISKKDFEKLLAQKSPLVEVKGEWIALQPADVKAQQIILNKSYPLELSVEDALRFSTGDISTV
AKLPITNFEAKGELANLINAINNNESIPIENPRGFKGQLRPYQQRGVWLSFLEKWGLGACLADD
MGLGKTPQLIGFLLHLRSEGMLDQPTLVICPTSVLNNWEREVQKFAPTLSTLIHHGDKRSKGKAFV

FIGURE 10 (continued)

48/96

KAVSKKNVITTSYSLIYRDIKSFEQVEWQGIVLDEAQNINPNQAKQSQAVRQISTQFRIALTGTPV
ENRLTELWSILDFLNPGLGTQQFFRRRFATPIEKYGDKESLQIMRSLVRPFILRRLKTDKTI IQD
LPEKQEMTIFCGLSSEQGKLYQQLVDNSLVAIEEKTGIERKGLILSLLLKLKQICNHPAHFLKQKS
LKTAEQSGKLLRLEEMLEELIEEGDHALIFTQFSEWGKLLQPYLQKKFQQDVLFLYGATRVRVQRQE
MIDRFQQDPNGPRIFILSLKAGGTGLNLTRANHVFHIDRWNPAVENQATDRAFRLGQKRNQVHK
FVCTGTLEEKINEMLESKQKLAEQTVDAGEQWLTELDTDQLRNLLLLDRDTI IDEQ

SEQ ID NO: 39, *Gloeobacter violaceus* PCC 7421 Glovi_SNF2 nucleic acid sequence

ATGGCTATCTTGCACGGTATCTGGGTTCACCAACCCCCCGGGCCGGGCTTTTCCTTTGGGGAGAA
ACCTGGAGGCAGGTTCGCAAAGCGGCGCAAGCGCTCCGAAGCACCCGCTCCGCATCCCTATGTCCAG
CAACCGGCCGAGTTGTCCCCCGCCTGGCTGCCCAGTTTCCCCAGATACCGCTCAGCTTGCTGGTA
CCCGAGACGCTTGCACTCCAGTTGCCCGCCACGGTCGAAAACGTGGTCTACTCCGCAAGCATTGCT
CCCGAGGGCAAGCTTTTGGAGTTGGAACCGTGGCTGGTGGAAAGTTTCTGGCTCGACGGTCACCAG
GCTTTTGAAGTGTGCTCGGGGTACCCCTGGGCGGCGGGGACGCATCGATTGGCGACGACCTGCGC
TTCTGGTCGCGAGTGCGCCCGCTGGGTGCTTGACTTGCTGGTGCGCGCCAAGTACCTGCCCGACCTG
GAGAGCGGCGACGGCCAGGAAATCCCCACAGCCCGCTGGGTGCCCCCTGCTCGACAGCGCCGTGAT
CAAGCCCGCCTCAAAGAATTTGCCGCCCGTTTGCCGGGCGCCTGCCGCGCCGCTACCCCCGAACCTA
TCTCCGCACCAGATTCTCAAGAGTTTCTGAGCGCCATGCTCGACGCGCGGGTGCGCACGCTGCTC
GCTTGCGAGCCTCCCGATCCGCGCACGCTGCCTGCCGGAGCGGTGCGCCCCCTGGCTTCTGGCCCTG
GCCCCATGCCAGCCCCAGCTCAAATCTCCGGACCCGGAGACGCCGGCTCTGGCGGAAGCCCTGGCC
ACCTGGCGCGCCCCCCTGAGCTATCAGGTTTCGCTCGCGCACCTGCTTCCGTCTGCAGCCGCCCGAG
GAGAGCCAGGGCGAGTGGAAGCTGCACCTTCTATTGCAAACAGGCGACGATCCCGATTGCTGATG
GCTGCCCAGCAAGTCTGGAGCAGCGCGGGTGAGCTGCAGGAGGTGTTTCTCGCGGGCTTGGGCCTC
GCCTCGCGTATCTTTGTGCCCGTCGAGCGGGGATTGCTCGTCCCCCAGCCACCTGCTGCACCATG
AGCACCGTCGAGGCGTTTTCAGTTTCTCAAAGCCGCCACCTGGCGGTTGCGCGACAGCGGCTTCGGG
GTGTTGTTGCCCGAGAGCCTCGCGGACGCGGGCAGCCTGCGCAACCGCCTGGGCCTCAAACCTCGAA
GCGAACGCGCCGGGGCGCAACGGTTCGGGCCTCGGCATGCAGAGCTTGCTCGCTTTTAAATGGGAG
CTGTCGCTCGCGGGCAAGACCCTGAGCCGCGCCGAGTTTCGACCGCCTCGCCGCTAGTTCTGAACCC
CTGGTCAAAGTCAACGACAACCTGGGTGCAATTGCGCCCCCAGGACGTGCGCGCCGCCACAGCTTT
TTGCAGTCGCGCAAAGATCAGGTTCGACTCTCGTTGGAGGATGTGCTGCGCCTCAACTTCGGCGAC
ACCCCCAAAATCGACGGTCTCCCCATCGTCAACTTCGACAGCTCCGGCCCCATTAGCAACTGCTG
GAGACCCTCACCGATCAGCGCAAACCTACCCCCATCGACGAACCGCCGGGGTTCAAGGGCACCTG
CGGCCCTATCAAAAAATTGGCGTCGGCTGGCTCGCCTTTTTTGAGAAGTGGGGCCTGGGTGCTTGC
CTAGCCGACGACATGGGACTCGGGAAGACCGTAGAGTTGATAGCATTTCTTTCTTTTCTCAAATCC
AAAAATGAGCTGGACGGCCCTATATTGTTAATTTGTCCGACTTCAGTGATGGGAAACTGGGAAAGA
GAAATAAAGAAATTTTCTCCTAGTTTATCTGTACATGTCCATCATGGGGCGCGGCGGCCGAAGGGG
CGCAATTTTGTGCGAGACGGCCCAGAAAAAGCAAATCATCGTCAGCAGCTACGCCCTGGTACAGCGC
GACAGCAAAGATCTCAAGCGCGTCGAATGGTTGGGCCTGGTGCTCGACGAAGCCCAGAACATCAAA
AACCCCGACGCCAAGCAGACCCAGTCGATTCGGGAACTGACAGCGCGCTTTTCGCATCGCCCTCACC
GGCACACCGGTCGAGAATCGCCTCGCGGAACTGTGGTCGATCCTCGATTTTCTCAATCCCGGCTAT
CTGGGGGCGCGCAACTTCTTTCAGCGCCGCTTCGCAGTTCCGATCGAAAAGTACGGGGATCGCTCC
TCGGCGAACGCCCTCAAAGCTCTGGTGACGCCGTTTATCCTGCGGCGGCTCAAATCCGACCCGCGAG
ATTATTCAAGATCTGCCCCGAGAAGCAGGAGACGAATGTCTTCTGTCCGCTCACACCCGAGCAGGCG
GCCCTCTACGAGCGGGTGGTGAACGAATCGCTCGCCAAGATCGAGCAGAGCACCGGCATCCAGCGG
CGCGGGACGGTGCTGGCCACCTTGGTCAAACCTCAAGCAGATCTGCAACCAACCCGAGCCACTACCTG
GGTGACGACGGACCGCTCGCCAACCGCTCGGGCAAACCTCAGCCGCCTGGGCGAGATGCTCGAAGAA
GTGCTCGCCGACGAGGAGCGGGCGCTGATTTTTTACCCAGTTCGCCGAGTGGGGCCACCTGCTGCAG

FIGURE 10 (continued)

49/96

GCGCACCTGAGCCGCCAGTTGGGTTCAGAAGTGTTTTTCTCTACGGCGGCACCAGCAAAAACCAG
CGCGAGGCGATGATCGAGCGCTTCCAGAGCGATCCGCAGGGGGCCGCGGATTTTTTATTCTTTTCGCTG
AAGGCAGGGGGTGTTCGGCCTCAACCTCACCCGCGCCAACCACGTCTTCCACTTCGACCGCTGGTGG
AACCCGGCGGTCGAGAATCAGGCCACCGACCGCGTCTTCCGCATCGGCCAAACCAAGAACGTACAA
GTCTACAAGTACGTGTGCACCGGCACGCTCGAAGAGCGCATCAACGCCCTGATCGAAAGCAAAAAG
GCCCTGGCTGAGCAGGTGGTGAGCGCCGGTGAGAACTGGCTGTTCGGATCTAAATACCGATCAACTG
CGGCAACTGTTGGTACTCGATCGCTCGGAGATTATCGACACGGAGGACACCGCGTGA

SEQ ID NO: 40, *Gloeobacter violaceus* PCC 7421 Glovi_SNF2 translated polypeptide

MAILHGIWVHQPPRAGLFLWGETWRQVAKRRKRSEAPAPHPYVQQPAELSPRLAAQFPQIPLSLLV
PETLALQLPATVENVVYSASIAPEGKLLLELPWLVEGFWLDGHQAFELLLGVPLGGGDASIGDDL
FWSQCARWVLDLLVRAKYLPDLESGDGQEIPTARWVPLLDSDAVDQARLKEFAARLPACRAATPEL
SPHQILKSFLSAMLDAVRVTLTLLACEPPDPRTLPAHAVRPWLLALAHQAQPLKSPDPETPALAEALA
TWRAPLSYQVRSRTCFLRLQPPEESQGEWKLHFLLLQTGDDPDSLMAAQVWSSAGELQEVFLAGLGL
ASRIFVPVERGLLVPQPTCCTMSTVEAFQFLKAATWRLRDSGFGVLLPESLADAGSLRNRLGLKLE
ANAPGRNGSGLGMQSLLAFKWELSLAGKTLRAEFDRLAASSEPLVKVNDNWVELRPQDVRAAHSF
LQSRKDQVGLSLEDVLRNLNFGDTPKIDGLPIVNFDSSGPIQQLLETITDQRKLTPIDEPPGFKGTL
RPYQKIGVGWLAFLQKWGLGACLADDMGLGKTVELIAFLLFLKSKNELDGPILLICPTSVGMNWER
EIKKFSPSLSVHVHVGARRPKGRNFVETAQKKQIIVSSYALVQRDSKDLKRVEWLGLVLDEAQNIK
NPDAKQTQSIRELTARFRIALTGTPVENRLAELWSILDFLNPGYLGARNFFQRRFAVPIEKYGDRS
SANALKALVQPFILRRLKSDPQIIQDLPEKQETNVFCPLTPEQAALYERVVNESLAKIEQSTGIQR
RGTVLATLVKLKQICNHPSHYLGDDGPLANRSGKLSRLGEMLEEVLADEERALIFTQFAEWGHLLO
AHLRQLGSEVFFLYGGTSKNQREAMIERFQSDPQGPRIFILSLKAGGVGLNLTRANHVHFHFDWW
NPAVENQATDRVFRIGQTKNVQVYKYVCTGTLEERINALIESKKALAEQVVSAGENWLSDLNTDQL
RQLLVLDREIIDTEDTA

SEQ ID NO: 41, *Lyngbya* sp. PCC 8106 Lyn_sp_SNF2 nucleic acid sequence

ATGGCAATTTTACACGGAAGTTGGCTCCAGCACCCCAAAAATTATTTGTTTATTTGGGGAGAAACC
TGGCGTCGCATTACACCCAATGAATTTAATCCGGCTGATGGTGTGTTTTGGGTTATCCTTTTGCTTTA
AGCCCTGTTGAATTGGAAGAGTGGTGCAGTGAAAAGCAGTTATCTATAGAGAGTAAAGTTGTCGTT
ACAGAAACTCTCGCCCTTCCCACTAAACTCTCCCCAAAAATAGGACTATATCCCCTTCAATCTACG
CCTCAAACCTGATTCTGAAACTGATTCTGAGTCGATCTGTCTTTATCCCTGGAAAATTGAAGGTATT
TGTCTCAACAGTACAGAAGCCTTTGACTTTTTTACAATCCCTTCTCTGGGAAACCTGACCACAGAA
AACTCATTTTATTGGCTCAGATTTACAGTTTTTGGTCTCATCTTTCCCGTTGGAGTTTACTTACTC
GCCCCGGAGTAAATTTTTTACCCAGTCTCACTTTTAACCCCTCAAAAGATCACTTTATCGCTGAATGG
AAACCTTTACTCGATAGTGCGACAGATCAAGCCAGATTAATTCGTTTTTCTAAACAAATACCCCTCT
GCTTGTCGGATCTATCAACTCTGGTCAAAAGAGGCTCAAAATCAATTTGAAAATTTAGCCCTAGAT
TTACCTCAAAATCCCCAAAACCTTAATTGATGATTTTTTTAACGGCAATTATTGATAGTCAAGTCAAG
AAAGTTGCAGAAGAAAGTGAAGAAAAAGCGATTACAAATCTAACCGCTATTCAACCGATTGTTTCAG
AGTTGGTTACACGCTTTAGCCAGTGAATCTAATCTAGCAAAATCCAAAAAATCTGAATCAAAAACC
CTAGAAAAAATTCTTTCCAATTGGACGGCTCCTCTTCAACAAACTCTCGCTGAACATAATTTGTTT
AGAACGGGATTTCTGACTCTCTCCTCCGGAATAATCAAAAAAATTGGACGCTAGATTATTGTTTA
CAAGCAATTGATGAACCCGAATTTTTTAGTGGATGCTCAAACCTATTGACTCATCCAGTCGAAGCC
TTTGTTTCACAATGGACGTATGATTAAACGTCCTCAAGAAACCCTCCTCAAAGGTTTAGGTTTAGCC
TCAAAACTATATCCTCTCCTAGAACCCAGTTTACAAGAAGCCCGTCCTCAAACCTTGCTTATTAACG
CCCCTACAAGCCTATGAATTTATTAAAAGTATTAATTGGCGGTTTACAGATAGCGGTTTAGGAGTG

FIGURE 10 (continued)

50/96

ATTTTACCCCCGAGTTTAGTCAGTCAAAATGGATGGGCGAACCGTTTAGGTTTAAGTGTTCAAGCG
GCGACATCAAAATCCAAACAAAATGTTAGCTTGGGATTAGATAGTCTGCTGAATTTTAAATGGGAA
TTGTCAATTGGGGGTCAAACCTTATCAAAAACAGAATTTAACCGTTTAGTCGCTCAAGAAAGTCCG
TTAGTTGAAATTAATGGCGAATGGGTGGAATTACGTCCTACTGATATTAAAGCCGCTAAAGCCTTC
TTTTTCGAGTCGCAAAGATCAACTTTCACTTACCCTTGAAGATGCTTTACGTTTATCGACGGGTGAC
TCGCAAATGGTGGAAAAGTTACCGATTGTTAACCTTTGAAGCGGGTGGAAAATTAGAAGAACTTCTC
AATACTTTAACGAATAACCGTTCGCTCGATGAGATCAAAACTCCTAGTAATTTTCAAGGAGAACTA
CGCCCCCTATCAAGCCCCGAGGGGTGAGTTGGTTAGCCTTTTTTAGAAGAATGGGGTTTAGGGGCTTGT
TTAGCTGATGATATGGGGCTAGGAAAAACCATAGAATTAATTGCTTTTCTCTTGTATTTGCAGGAA
AAAGAAACCTTAGACGCTCCTGTTTTACTGGTTTGTCCGACATCAGTTTTAGGAAACTGGGAACGA
GAAGTTAAACGATTTAGTCCGAGTTTAAAAGTTACTGTTTCATCACGGGGATAAACGCCAGAAAGGG
AAAAACTTTGCTCAATTTGCCCAGAAATATAATTTAATTATTACCAGTTATCCGTTAACTTTTCGA
GATGAGAAAGAACTCAAAACGGTAAATTGGAAAGGATTAGTTTTAGACGAAGCTCAAAATATTAAA
AATCCCGAGGCTAAACAATCAAAAACGGTGAGAAATCTACAGGCGAGTTTTTAAAATTGCTCTGACT
GGAACACCTGTGCGAAAACCGTCTGTCTGAATTATGGTCAATTATGGATTTTCTCAACCCAGGTTAT
TTAGGACAGCGACAATTTTTTTCAGCGAAGATTTGCTATTCCGATTGAAAAATACGGCGATACAGAC
TCCTTAAAAACATTGCGATCTTTGGTTCAACCGTTTATTTTACGGCGCTTAAAAACAGATAGAGAG
ATTATCCAAGACTTACCCGAAAAACAGGAAAATACGATCTTTTGTCTCTGTCTACAGAACAAGCA
ACGCTTTTATCAAAAGATTGTTGATCAGTCTTTAGCTGACATAGACTCAGCCGCAGGAATTCACGT
CGAGGGATGATTTTAGCGTTGTTAGTGAAATTAAAACAGGTTTGTAATCATCCCATTTTATTGAAT
GGAAAAGCGACAAAAACTGGAAAGAAAAAGGTCGAGACTCAGGGTTTAAGCCTGCAAAGTTCAGGG
AAGTTACAACGCTTCAAAGAAATGCTGGAAGAATTGTTGTCAGAAGGAGATCGCGCCATTGTATTT
ACCCAGTTTGCAGAATGGGGAAAAGTTTTACAACCTTATTTAGAACAGCAATTAAACCGAGAGGTA
TTATTTTTTGTATGGCGCAACTCGTAAAAATAAACGAGAAGAAATGATTGATCGTTTTTCAACAAGAT
CCTCAAGGGCCACCGATTTTTTATTCTATCTTTAAAAGCGGGAGGTGTGGGTTTAAATTTGACTCGT
GCTAATCATGTTTTTCACTTTGATCGTTGGTGAACCCCTGCGGTTGAAAATCAAGCAACAGATCGG
GTGTTTAGAATTGGTCAAACGCGCAATGTTTCAGGTTTCATAAGTTTGTCTGTACCGGAACGTTGGAA
GAAAAAATCCATGATTTAATTGAAAGTAAAAAAGTGTTGGCTGAACAAGTTGTGGGTTTCAGGAGAA
AATTGGTTAACTGAATTGGATACGGATCAACTCAGAACTTACTCATTATTGACCGAAATGCGGTG
ATTGATGAAGAAGAATAA

SEQ ID NO: 42, *Lyngbya* sp. PCC 8106 *Lyn_sp_SNF2* translated polypeptide

MAILHGSWLQHPKNYLFIWGETWRRITPNEFNPADGVLGYPFALSPVELEKWCSEKQLSIESKVVV
TETLALPTKLSPKIGLYPLQSTPQTDSETDSESI CLYPWKIEGICLNSTEAFDFLQSLPLGNLTTE
NSFIGSDLQFWSHL SRWSLDLLARSKFLPSLTFNPSKDHFI AEWKPLLD SATDQARLIRFSKQIPS
ACRIYQLWSKEAQNQFENLALDLPQNPQN LIDDFLTAIIDSQVKKVAEESEKKAITNLTAIQPIVQ
SWLHALASESNLAKSKKSESKTLEKILSNWTAPLQQTLAEHNLFR TGFR LSP PENNQKNWTL DYCL
QAI DEPEFLVDAQTIWTHPVEAFVHNGRMIKRPQETLLKGLGLASKLYPLLEPSLQEARPQTCLLT
PLQAYEFIKSINWRFTDSGLGVILPPSLVSQNGWANRLGLSVQAATSKSKQNVSLGLDSL LNFKWE
LSIGGQTL SKTEFNRLVAQESPLVEINGEWVELRPTDIKAAKAFFSSRKDQLSLTLEDALRLSTGD
SQMVEKLPIVNFEAGGKLEELLNTLTNNRSLDEIKTPSNFQGELRPYQARGVSWLAFLEEWGLGAC
LADDMGLGKTIELIAFLLYLQEKETLDAPVLLVCPTSVLGNWEREVKRFSPSLKVTVHHGDKRQKG
KNFAQFAQKYNLIITSYPLTFRDEKELKTVNWKGLVLDEAQNIKNPEAKQSKTVRN LQASF KIALT
GTPVENRLSELWSIMDFLNP GYLGRQFFQRRFAIPIEKYGD TDSLKT LRS LVQPFILRRLKTDRE
IIQDLPEKQENTIFCSLSTEQATLYQKIVDQSLADIDSAAGIQRRGMILALLVKLKQVCNHPILLN
GKATKTGKKKVETQGLSLQSSGKLQRFKEMLEELLSEGDR AIVFTQFAEWGKVLQPYLEQQLNREV
LFLYGATRKNKREEMIDRFQQDPQGPPIFILSLKAGGVGLNLTRANHVHFDRWWNP AVENQATDR
VFRIGQTRNVQVHKFVCTGTLEEKIHD LIESKKVLAEQVVGSGENWLTELD TDQLRNLLIIDRNAV
IDEEE

FIGURE 10 (continued)

51/96

**SEQ ID NO: 43, Methanosarcina acetivorans C2A Metac_C2A_SNF2
nucleic acid sequence**

ATGATAATTTTGCATGCAGGAAGAGTCGGAAAACAGTTCTTTCTGTGGGGCGAAAGCCCGGCTGAA
AATGAAACTCCGCCTGTCCGGCGCGGGAGAAAGCCTAAGAAGCCGGTTGCAAAACCTTATCCTTAC
GATTCGGGTGTTGAAAACCTGTCTTCTGCTCTTGAGCTGCTGCTGGGCAGTACTGGCCGGAAAAAG
GCAGAGGAAATCAATGTCTGGATCCCGACAGCAGGCTGGAATCCAATCCCCTCCAGTCCTCTCGTT
GCTGAAATTCCGGCTTCGAAAGCAGAACTTTCCCTAGCTCCCTGGACTGTTACGCATATCCTCTG
GAAGCTGAAGAAGCTATTGTTCTCCTCTGCGCCTGTATGGGAAAAAAGGTTCTTGCTCCCGGCATA
ATCTCGGGAAATGATCTTCTCTGGTGGGCGGATGCCCTGAAATTTGCAGGCTCGCTGGTAGCAGGA
CAGAAATACCTGCCTGGCGTCAGGGGCGGGGAAGGAGAGTACAAGGCTTTCTGGGAACCCGTATTT
TCCGGAGAAGATGCGGGGGAGCTGGCAAGACTTGCAAAGCAAATGCCTCCGGCTGCAAAGGCTCTT
GCTCTTGAAACCTCTTCCGTGCAGCCGGAATACTTGCTGCTGTAGCGGCAAGGCAGTTTATCGAA
GAGGCTCTTGACTGGATAGTCCGGTCCGAGATCGGGGAAAAAGAGCTTGCAAAGAGGGCGCGTAAA
AGAAAATCCTTTGATAGCGTCCATGACGCCTGGGTTTCCGCTCTTAAAAGCCCTGACGGGTTGATC
CACGGAGAAGAAAAAGAACTCCTGCAGCTTGCGTTCCGGACCCGTGAATGGCAGCGCCCCCTTACT
GTACTTACAACCTTCTCCCTTCAGGTTCTGTTTCCGGCTTGAAGAGCCAGCTGCGGAAGAAGAACTC
GAAGAAACCGAGGAATCCGAAGCCGGA AAAATGGATACTAAAAAAGGCAGGAAAGGGATAGCTGAC
ATAGAAGTTCCCGAAGAACTCTGGTACGTCCGCTATATGCTTCAGTCCTACGAAGACCCAAGCCTT
CTGATTCCTGTAAAAGAGGCCTGGAAACCAAAGAAGGGCAGCCCGTTGAAAAGATATGATGTAAAA
AACATTCGCCAATTTCTGTTATCTTCCCTTGGACAGGCTGCTGGCATCAGTGCAGGAATTGCTTCC
AGCCTTGAAGCTCCCAACCCGTCCGGATATTCCCTTGATACGAAAGAAGCTTACCGCTTCCTGACT
GAAAGTGCAGCGGATTTAAGCCAGGCGGGCTTCGGGTTACTTCTCCCCGGCTGGTGGACCCGTAAA
GGTACAAAGACCCACTTAAAAGCCCAGGCTAATGTTAAGGGCAAGAAGTTGAAGGCCGGATACGGG
CTTACACTCGATAAAAATCGTCAGCTTTGACTGGGAAATTGCCCTTGGAGACCGTGCACTCACAGTC
AGGGAACTGCAGGCTCTTGCAAAGCTCAAAGCTCCGCTTGTGAAATTCCGCGGGCAGTGGGTGAG
GTCAACGATGCGGAAATCCGGGCTGCCCTTGAGTTCTGGAAGAAAAACCCCCACGGGGAAGCAAGT
CTGCGCGAAGTTCTAAAAGTGGCTGTGGGAGTCTCCGAAAAAGCCGATGGTGTAGACGTTGAAGGG
CTTAATGCAGCCGGCTGGATCGAAGAATTAATCCGCCGCCTGAAGGACAAAACCGGGTTTGAAGAA
CTTCCGGCTCCTGACGGTTTTTTCAGGCACCCTCAGGCCCTACCAGTTCAGAGGTTACTCCTGGCTG
GCTTTCCTGAGGCAGTGGGGCATAGGAGCCTGCCTTGCAGACGACATGGGGCTTGGTAAAACCATC
CAGACCCTTGCCCTTATCCAGCACGACCTGGAACAGGTTAAAGGGCAGGTTGAAGAAAAGGTTATA
GAAAATGCTGAAGAAAAAGTTGAAGGACTTAAAGCTGCAAACCGGTTCTTCTGGTCTGTCCGACC
TCTGTCAACAACCTGGA AAAAAGAGGGCGGCTCGCTTTACCCCGGAACTTTCCGGTAATGGTCCAC
CACGGGACCAGCCGGAAAAAGGAAGAGGAATTCAAAAAGGAAGCCACGAATCATTCTATTGTCGTC
TCAAGCTACGGGCTTTTGCAGCGGGATCTTAAGTTTTTAAAAGGGGTTTCCTGGGCCGGAGTGGTA
CTTGACGAAGCCCAGAATATCAAAAACCCGGAAACCAAACAGGCAAAGGCAGCCAGAGCTCTTGAA
GCCGATTACCGCATAGCTCTTACGGGGACTCCGGTTGAAAACAACGTGGGAGACCTCTGGTCTATC
ATGGAGTTTTTTAAACCCCGGCTTCCTAGGCAACCAGGCAGGTTTCAAGCGGAATTTCTTTATTCCC
ATTCAGGCCGAAAGGGATCAGGAAGCTGCAAGGAGGTTAAAAGAAATTACGGGCCCCCTTTATCCTG
CGCCGTCTGAAGACCGATACTTCGATTATCTCCGACCTGCCGGAAAAGATGGAAATGAAAACCTAT
TGTACGCTGACAAAAGAACAGGCTTCCCTCTATGCCGCAGTCCTCGAAGACATCGAAGAGACGATG
GAAGAGGCTGAAGAAGGCATCCAGAGAAAAGGTATAATCCTGTCCGCCCTTACCAGGCTCAAACAG
GTCTGCAACCATCCGGCGCAGTTTTTGAAGGATAACTCTGCTGTACCCGGCAGGTCAGGAAAACCTT
GCAAGGCTTACCGAAATGCTGGATGTAATCCTGGAAAATGGGGAAAAAGCCCTTGTGTTACCCAG
TTTGCGGAGATGGGAAAAATGCTAAAAGAACACCTGCAGGCAAGTTTTGGCTGTGAAGTCCTTTTC
CTGCACGGCGGGGTCCCCAGAAAGCAGAGGGATCGGATGCTTGAGCGTTTTCCAGGAGGGAAAAAGAA
TACCTCCCTATCTTTGTCCTCTCCCTTAAAGCTGGAGGCACGGGGCTTAACCTTACAGGAGCGAAC
CACGTTTTCCATTTTGACCGCTGGTGGAAACCCTGCTGTTGAAAACCAGGCTACGGACAGGGCTTTC

FIGURE 10 (continued)

52/96

CGTATAGGCCAGACGAAAAATGTAGAGGTGCATAAGTTCATCTGTGCGGGTACGCTTGAAGAAAA
ATCGATGAGATTATCGAGCGCAAAGTGCAGGTTGCAGAGAACGTTGTGCGAACAGGTGAAGGTTGG
CTGACAGAACTTTCCAACGAGGAATTGAAGGATATTCTTGCTCTCCGAGAAGAAGCGGTAGGTGAA
TAA

SEQ ID NO: 44, Methanosarcina acetivorans C2A Metac_C2A_SNF2 translated polypeptide

MIILHAGRVGKQFFLWGESPAENETPPVRRGRKPKKPVAKPYPYDSGVENLSSALELLLGSTGRKK
AEEINVWIPTAGWNPI PSSPLVAEIPASKAELSLAPWTVHAYPLEAEEAIVLLCACMGKKVLAPGI
ISGNDLLWWADALKFAGSLVAGQKYLPGVRGGEYKAFWEPVFSGEDAGELARLAKQMPPAKAL
ALETSSVQPEILAAVAARQFIEEALDWIVRSEIGEKELAKEARKRSFDSVHDAWVSALKSPDGLI
HGEEKELLQLAFRTREWQRPLTVLTTSPFRFCFRLEEPAAEEEELEETEESEAGKMDTKKGRKGIAD
IEVPEELWYVRYMLQSYEDPSLLIPVKEAWKPKKGSPLKRYDVKNIRQFLLSSLGQAAGISAGIAS
SLEAPNPSGYSLDTKEAYRFLTESAADLSQAGFGLLLPGWWTRKGTKTHLKAQANVKGKKLKAGYG
LTLDKIVSFDWEIALGDRALTVRELQALAKLKAPLVKFRGQWVEVND AEIRAALEFWKKNPHEAS
LREVLKLAVGVSEKADGVDVEGLNAAGWIEELIRRLKDKTGFEELPAPDGFSGTLRPYQFRGYSWL
AFLRQWGIGACLADDMGLGKTIQTLALIQHDLEQVKGQVEEKVIENAEKVEGLKAAKPVLLVCPT
SVINNWKKEAARFTPELSVMVHHGTSRKKEEEFKKEATNHSIVVSSYGLLQRDLKFLKGVSWAGVV
LDEAQNIKNPETKQAKAARALEADYRIALTGT PVENNVGDLWSIMEFLNPGFLGNQAGFKRNFFIP
IQAERDQEAARRLKEITGPFILRRLKTDTSIIISDLPEKMEMKTYCTLTKEQASLYAAVLEDIEETM
EEAEEGIQRKGIILSALTRLKQVCNHPAQFLKDNSAVPGRSGKLARLTEMLDVILENGEKALVFTQ
FAEMGKMLKEHLQASFGEVLFHGGVPRKQRDRMLERFQEGKEYLP I FVLSLKAGGTGLNLTGAN
HVFHFDRWWNPAVENQATDRAFRIGQTKNVEVHKFICAGTLEEKIDEIIERKVQVAENVVGTGEGW
LTELSNEELKDILALREEAVGE

SEQ ID NO: 45, Methanospirillum hungatei JF-1 Methu_JF-1_SNF2 nucleic acid sequence

GTGACCGCGAAACGACCAGCACCAATCCACGATAAAGAAGAAGAGACCATAACCGATACTTCGCTT
CCGGTCTTTCATGCCCTGATTTACCCGGCCGTTGAAGGGGTAGCGATATGTGCCGAATATATAACT
GATAAACCTGCACCGGTCAGGAAAAAAGGCTACGCAAAGGATAAACCTGGCGAATATCCATATTCC
CTGGATCATAACCGCCCTTAAACGCTCATAGAGA ACTGTTTTGGAGCATATGATGACCTGAAGGCT
ACCAGATGGATTATCTATCTCCCCGCTGAAGAAACGGTTCCTCCTTCTCAGTTCTCATCAAAA
AAGAAGCCATCACCAAAGGAGAAAAAACTCCCCCTTGTTCCGATGTATATCCCCGTTCTTCTCTGC
CCGTATGAAACCTTTTTTCAAATCTGGAAAGCCGCTCAGAATACAGATAAAAATTATATTGCTGGC
GATTCCTTCCAGTACATCTCCATTCTGATGGAGAGTACCGTCCGGCTCATACAAAACGGACGGTTC
AAACCATCTCTAGAACGGACCTTTGCCGGATATCATGCCGTATGGGTACCTGCCCTTTCTCCTCAG
GATATGGAATGGGTATCAGATTTTTCAAGCCGGATGCCAACGGTCTGCAAGTACGCTATCCCCCGG
GTCGCAAAAGATCCCTACATTTATAAACCTGAGACCAGATTAGAGAAATTCATCGTTGAGATGATG
CGGGTGATCATCCGTACTGCCCTTGGTGGTTATACTGAAAGAAGAGACAGATCCCTTTTATGAA
CCCTCAGAAAACGAGATGCAGTTCATGACTGACCTTCTCGGGGTAACCGACCCAATAAGGAACAAA
GGATTTGAGAGAACTTTCTTACGGGCGATGCAGGACTGGCTGACCTTCTCAAGTTCAGGACGGTTT
GCTCCCTTTGAGTTCTGCATGATCATAAAAGATCCACCAGAAGGACAGACAGAACCATGGGATTTT
ACTCTCGCGGTCAGATCAGAGGCAGAACCATCTCTTCTCATCCCGGCAGAAATAATCTGGGAATTG
CCTGATCACCAGAGCGGGCTCTTCCCCCAGGCAGCCTATCTCAAACATATCCTCCTTGCTGGTATC
GGGCTCTTGACCTCATCATCATCGGCATTATGGCGTCCCCTGTCCGGATCGAAACCCACCGGGGGA
AGTATGACCTGAAAGAGGCTGCAACGTTCTTGGGTTCAGACCTCGCAAGAGCCAGGAGGAAGGGA
GTAACGGTGCTCCTGCCAGACTGGTGGACTGATACGACCTATACACCACGGGTTGAAATCCATGCA
AGGCGGCGGGATCCCACCCATACGCAGACACGGATAGGACTGCAGGAACTCCTTTCTTTTGATTAC

FIGURE 10 (continued)

53/96

CGGATTGCAATCGGTGATGAGTCATTTTCACCGGATGAGTTCTGGGAAAAGGTAAAAGAAAAGGCT
CCCTTTATCTGGCTGGGGAACCGGTGGATATCCTTTCATCCGGATGCGATACAACATGCCCTGGAT
TCTTTCAGCAGGCATCAGAGCAAAGGAGGGGATACAATAGGAGATCTGCTCCGGCTCTCCCTGAAA
AAAATGGAGGATTCCGCGGTACCGGTATCGATTTCATGCAAAAGATGACTGGGTTGCGGATCTTCTG
GATTTTTTTCAGGACCGAAACAAATCAGGCAGTTCCAGTCCCAAAGAAATTTAAAGGGGATACTCAGG
CCATACCAGGAAGAGGGGTTCTCCTTCCTTTGTCAATGTACCAGAAGGGGCTTTGGAGCCTGCCTT
GCAGATGACATGGGGCTTGGAAAACTCCCCAGACACTTGCATGGCTGGTCTATCTCAAGGAGAAA
GAAAAACCCACGACTCCGTCCCTCCTTATATGCCCGATGTCGGTTGTTGGGAACTGGGAGCGGGAG
ATACAGCGGTTTGCGCCATCACTCCGTTTCATGGGTGCATCATGGGACTGACCGATGCAAAGGCGAT
GATTTTGTGAGACATGTCGGTTCATATGACCTGGTCCTGACCACCTATCATCTGGCAGCACGGGAC
GTAGACCACCTCAAAACCGTTCCTTGGTCTGCAATCATTTCTTGACGAGGCACAAAATATCAAGAAC
CTCCATGCAAACAGACCGTAGCAGTCAAATCTCTCACCGGTGAGAGACGGGTTGCTCTGACCGGA
ACCCCGGTGGAGAACCGGTTACTCGAACTCTGGTCTATCATGGACTTTTTTAAATCCAGGATACCTT
GGTTCACAGAGTGCATTTACAAACCGCTATTCCCGCCCCGATTGAGCAGGAAAAAAATACGGAAGT
ATACAGGAATTAAGGTCCCTCATCCGTCCGTTCCCTGCTCAGGCGGATGAAAACAGACAAGCATGTT
ATCGATGATCTTCCGGAAAAGATGGAGAACCGGGTATATTGCACCCTCACACCCGAACAGGCAACC
TTATATCAGGCTGTTGTGCTTGATATGGCAAAGAACCTTGATAAAGTGGAGGGTATTGCCAGGAAA
GGGGCAATCCTTGCTGCGATCACACGACTGAAACAGATCTGTAACCATCCGGGACGTGTTGGCAGG
GATAAAACAATAAAGGCTGAGCGGTCCGGGAAGGTGAGCCGGCTGCTTGAGATGATTGAGGAGATC
ACTTCCGAAGGGGACTCAGCACTCATATTTCAGTCAGTATGCAACATTTGCTGAGGAACTGGCAGGG
ATGATAGAGAAACAGGGAGATACGCCCCGTTCTTCTCCTGACCGGGTCAACACCACGGAAAAAACGG
GAACAGATGATAGAGGAGTTTCAGGCCTCAACCACCCCGATAATCTTTGTTATTTCTCTGAAAGCC
GGGGGAACGGGTCTGAACCTGACGAAAGCGACTCATGTGTTTCATGTAGACCGGTGGTGGAAATCCG
GCGGTTGAAGACCAGGCTACTGACCGGACGTACCGGATCGGACAAAAGAGAAATGTCCAAGTTCAC
CTGATGATAACCGCCGGAACCTGGAGGAACGGATAGATCTGATAAACCAGGAGAAACGGACGCTT
GCAAAGGAAGTCCTTGACACAGAGTGATGAGTATCTGACAAATCTCTCAACAAAAGAACTTCTGGAG
ATTGTATCACTTCGTGACAGTCTCTTTCGCGGGGAGGATGCATGA

**SEQ ID NO: 46, Methanospirillum hungatei JF-1 Methu_JF-1_SNF2
translated polypeptide**

VTAKRPAPIHDKEEETIPDTSLPVFHALIYPAVEGVAICA EYITDKPAPVRKKGYAKDKPGEYPYS
LDHTALKTLIENCFGAYDDLKATRWIIYLP AEETVPPSSQFSSKKKPSPEKKLPLVPMYIPVLLC
PYETFFQIWKAAQNTDKNYIAGDSFQYISILMESTVRLIQNGRFKPSLERTFAGYHAVWVPALSPQ
DMEWVSDFSSRMPTVCKYAI PRVAKDPYIYKPETRLEKFIVEMMRV IIR TALGGYTLKEETDPFYE
PSENMQFMTDLLGVTDPIRNKG FERTFLRAMQDWLTFSSSGRFAPFEFCMI IKDPPEGQTEPWDF
TLAVRSEAEP SLLIPAEIIWELPDHQSGLFPPQAAYLK HILLAGIGLLTSSSSALWRPLSGSKPTGG
SMTLKEAATFLGSDLARARRKGVTVLLPDWWTDTTYT PRVEIHARRRDPHTHTQTRIGLQELLSFDY
RIAIGDESFS PDEFWEKVKEKAPFIWLGNRWISFHPDAIQHALDSFSRHQSKGGDTIGDLLRLSLK
KMEDSAVPVSIHAKDDWVADLLDFFRTETNQAVPVPKKFKGILRPYQEEGFSFLCQCTRGRFGACL
ADDMGLGKTPQTLAWLVYLKEKEKPTTPSLLICPMSVVG NWEREIQRFAPSLRSWVHHGTDRCCKGD
DFVRHVGSYDLVLT TYHLAARDVDHLKTVPSAIILDEAQNIKNLHANQTVAVKSLTGERRVALTG
TPVENRLLELWSIMDFLNP GYLGSAFTNRYSRPIEQEKNT ELIQELRSLIRPFLLRMKTDKHV
IDDLPEKMENRVYCTLTPEQATLYQAVVLDMAKNLDKVEGIARKGAILAAITRLKQICNHPGRVGR
DKTIKAERSGKVSRLLEMIEEITSEGDSALIFSQYATFAEELAGMIEKQGDTPVLLLTGSTPRKKR
EQMIEEFQASTTPIIFVISLKAGGTGLNLTKATHVFHVDRWWNPAVEDQATDRTYRIGQKRNQVH
LMITAGTLEERIDLINQEKRTLAK EVLAQSDEYLTNLSTKELLEIVSLRDSLFRGEDA

FIGURE 10 (continued)

54/96

SEQ ID NO: 47, *Methanosarcina mazei* Goel Metma_Go1_SNF2 nucleic acid sequence

ATGATAATTCTTCATGCAGGAAGAGTTGGAAAACAGTTCTTCTTATGGGGTGAAAGCCCCGGCAGAA
AATGAAACTCCGGTTGTTTCGGCGCGGGAGAAAGCCTAAAACCCCTATCGTAAAACCTTACCCTTAC
GATTCGGGCTTTGAAAACCTGTCTTCTGCCCTTGAGCTGCTGCTGGGCAGTACTGACCGGAAAAAG
GCGGAGAAAATCAACGTCTGGACCCCAACTATCGGAGGGAATCCTGTCCCTTCCAGCCCTCTTGT
GCTGAAATTTTCGGATTTCGAAAGCAGAACCTGCACTGGCTCCCTGTACTGTTTCACGCATATCCTCTG
GAAGCTGAAGAAGCTATTGTTCTCCTCTGCACCTGTATGGAAAAAAGGTTCTGGCTCCCGGTATC
ATCTCGGGAAATGACCTTCTCTGGTGGGCAGATGCCCTGAAATTTGCAGGCTCGCTGGTAGCAGGG
CAGAAATATTTGCCTGGCGTCAGGGGCGGGGAAGGAGAGTACAGGGCTTTCTGGGAACCCGTATTT
TCCGGCGAAGATGCCGGAAAGCTGGCAAACCTTGCAAAGCAAATGCCTCCTGCTGCAAGGGCTCTT
GCTCCTGAAGCCTCTTCCATGCCGCCGGAATGCCTGCTGCTTTAGCGGCAAAGCAGTTTATTGAA
GACTCTCTCGACTGGATAGTCCGGTCCGAGATCGGGGAAAAAAGCTTGCAAAGAGACGCGCAAA
AGAAAATCCTTTGATAGCGTCCATGATGCCTGGGTTTCTGCTCTTAGAAGCCCTGAAGGGCTGATC
TATGGAGACGAAAACGAACTTCTGCAGCTTGCGGGCCCGGACCCGCGAATGGCAGCGCCCACTCACC
ATCCTTACCACTTCTCCTTTTCAGGTTCTGTTTCCGTCTTGAAGAACCGGCTTTAGAAGAAGAGATC
GAAGAACTGAAGAAACCGAAGAAATAGAAGAAATGAAGCCGGGAAAAGAGATACTAAAAAAGGC
AGGGAAGGGATAGCTGATATAGAAGTTCCCGAAGGGCTCTGGTACGTCCGTTATATGCTTCAGTCC
TACGAAGACCCGAGCCTTCTGATCCCTGTAAAAGAAGCCTGGAAGCCAAAAAAGGCAGCCCGTTG
AAAAAATACGATGTGAAAAACATTCGCCAATTCCTGTTATCTTCCCTTGGACAGGCTTCCAGTATA
AGTGCAGGAATTGCTTCGAGTCTTGAAGCTCCCAACCCATCTGGATATTCCCTTGATACTAAAGAG
GCTTACCGCTTTCTGACTGAAAGTGCAGCGAATTTAAGTCAGGCCGGTTTCGGGGTACTTCTCCCT
GGCTGGTGGACCCGTAAAGGTACAAAGACACACTTAAAAGCCCAGGCTAATGTTAAGGGCAAGAAG
AAGTTGCAGGCCGGATACGGGCTTACACTCGATGAAATCGTCAGCTTTGACTGGGAAATCGCCCTT
GGAGACAGGGTACTGACAGTCAGAGAACTGCAGGCTCTTGCAAAGCTTAAAGCTCCGCTTGTGAAA
TTCCGCGGGGCAGTGGGTTGAGGTAAACGATGCGGAAATCAGGGCTGCCCTTGAGTTCTGGAAGAAA
AATCCCAACGGTGAAGCAAGTCTGCGTGAAGTTCTAAAACCTGGCAGTGGGAGTTTCCGAAAAAGCC
GATGGTGTGAACGTTGAAGGGCTCAATGCAACCGGATGGATTGGAGAATTAATCAGCCGCTTAAAA
GACAAAACCGGGTTTGAAGAACTTCCTGCTCCCAACGGCTTTTCAGGCACCCCTTCGGCCATATCAG
TTCAGAGGTTACTCCTGGCTGGCTTTTCTGAGGCAGTGGGGTATAGGAGCCTGCCTTGCAGACGAT
ATGGGGCTTGGTAAAACCGTCCAGACTCTTGCTCTTATTCAGCACGATCTGGAACAGGCTAAAGAG
AAAGCTGAAGAAAAGATTGAAGAACCGGCTGAAGAAAAGATTGAAGAAAAGTTGACGGACGTAAG
GCCCCAAAACCTGTTCTTCTGGTTTGTCTTACCTCTGTTATCAACAACCTGGAAAAAAGAGGCTTCC
CGCTTTACGCCAGAACTTTCGGTAATGGTCCACCACGGGACCAGCCGGAAAAAAGGAAGAGGAATTC
AAGAAGGAAGCCATGAATCATGCTATTGTCTCATCTCAAGCTATGGCCTTGTGCAGCGGGATCTTAAA
TTTTTAAAAGAGGTTTCAATTGGGCAGGAGTTGTACTTGACGAAGCCCAGAACATCAAAAACCCGGAA
ACCAAACAGGCAAAGGCAGCCAGGGCTCTTGAATCCGATTACCGCTTAGCTCTTACAGGGACTCCG
GTTGAAAATAACGTGGGAGACCTCTGGTCCATAATGGAGTTTTTTAAACCCCGGCTTCCTCGGAAGT
CAGGCGGGTTTCAAGCGGAATTTCTTTATCCCCATTGAGGCAGAAAGGGATCAGGAGGCTGCAAGG
AGGCTGAAAGAAATTACAGGTCCCTTCATCCTTCGCCGTTTGAAGACTGACACTTCGATTATCTCC
GACCTGCCGGAAAAAATGGAGATGAAGACCTATTGTACGCTGACAAAAGAACAGGCCTCCCTCTAT
GCTGCAGTCCTTGAAGACATCAGAGAAGCGATTGAAGGAGCCGAAGAAGGCATCCAGAGGAAAGGT
ATAATCCTGTCTGCCCTTTCCAGGCTCAAGCAGGTCTGCAACCACCCTGCGCAGTTTTTTGAAGGAC
AACTCCACTATCCCCGGCAGGTCCGGAAAACCTCGCAAGGCTTACCGAAATGCTGGATGTAGTCCTG
GAAAACGGGGAAAAAAGCCCTTGTTTTTACCCAGTTTGCAGGAGATGGGCAAAATGGTGAAAGAACAC
CTGCAAGCAAGCTTTGGCTGTGAAGTCCTTTTCTGACGCGCGGGGTCCCCAGGAAGCAGAGAGAC
CGGATGCTTGAGAGGTTCCAGGAAGGAAAAGAATACCTCCCTATTTTTTGTCTCTCCCTTAAAGCC
GGCGGCACGGGGCTTAACCTCACAGGGGCAAACCACGTTTTTCCACTTTGATCGCTGGTGGAACCCG

FIGURE 10 (continued)

55/96

GCTGTTGAAAACCAGGCTACAGACAGGGCATTCCGTATAGGCCAGAAGAAAAACGTTGAGGTCCAT
AAATTCATCTGCGCAGGTACGCTTGAAGAAAAAATCGATGAGATTATCGAACGCAAAGTGCAGGTC
GCAGAGAACGTTGTTGGGACAGGTGAAGACTGGCTGACAGAGCTTTCCAACGATGAACTGAAGGAT
ATTCTTGCTCTTAGAGAAGAAGCGGTAGGTGAATAA

**SEQ ID NO: 48, Methanosarcina mazei Goel Metma_Goel_SNF2
translated polypeptide**

MIILHAGRVGKQFFLWGESPAENETPVVRRGRKPKTPIVKPYPYDSGFENLSSALELLLGSTDRKK
AEKINVWTPITIGGNPVPSSPLVAEISDSKAEPALAPCTVHAYPLEAEEAIVLLCTCMEKKVLAPGI
ISGNDLLWWADALKFAGSLVAGQKYLPGVRGGEYRAFWEPVFSGEDAGKLAKLAKQMPPAARAL
APEASSMPPEMPAALAAKQFIEDSLDWIVRSEIGEKKLAKETRKRKSFDSVHDAWVSALRSPEGLI
YGDENELLQLAARTREWQRPLTILTTSPFRFCFRLEEPALEEEIEETEETEEIEENEAGKRDTKKG
REGIADIEVPEGLWYVRYMLQSYEDPSLLIPVKEAWKPKKGSPLKKYDVKNIRQFLLSSLGQASSI
SAGIASSLEAPNPSGYSLDTKEAYRFLTESAANLSQAGFGVLLPGWWTRKGTKTHLKAQANVKGKK
KLQAGYGLTLDEIVSFDWEIALGDRVLTVRELQALAKLKAPLVKFRGQWVEVNDAEIRAALEFWKK
NPNGEASLREVLKLAVGVSEKADGVNVEGLNATGWIGELISRLKDKTGFEELPAPNGFSGTLRPYQ
FRGYSWLAFLRQWGIGACLADDMGLGKTVQTLALIQHDLEQAKEKAEKIEEPAEEKIEEKVDGRK
APKPVLLVCPTSVINNWKKEASRFTPELSVMVHHGTSRKKEEEFKKEAMNHAI VISSYGLVQORDLK
FLKEVHWAGVVLD EAQNIKNPETKQAKAARALESDYRLALTGTPVENNVGDLWSIMEFLNPGFLGS
QAGFKRNFFIPIQAERDQEAARLKEITGPFILRRLKTDTSIIISDLPEKMEMKTYCTLTKEQASLY
AAVLEDIREAIEGAEEGIQRKGIILSALSRLKQVCNHPAQFLKDNSTIPGRSGKLARLTEM LDVVL
ENGEKALVFTQFAEMGKMVKEHLQASFGCEVLFLHGGVPRKQQRDRMLERFQEGKEYLP I FVLSLKA
GGTGLNLTGANHVFHFDRWWNPAVENQATDRAFRIGQKKNVEVHKFICAGTLEEKIDEIIERKVQV
AENVVGTGEDWLTELSNDELKDI LALREEAVGE

**SEQ ID NO: 49, Mycobacterium bovis BCG Pasteur 1173P2 Mycbo_SNF2
nucleic acid sequence**

ATGCTGGTTTTTGCACGGCTTCTGGTCCAAC TCCGGCGGGATGCGGCTGTGGGCGGAGGACTCCGAT
CTGCTGGTGAAGAGCCCGAGTCAGGCGCTGCGCTCCGCGCGGCCACACCCGTTTCGCGGCGCCCGCT
GACCTGATCGCCGGCATAACATCCGGGCAAACCCGCAACCGCCGTTTTTGCTGTTGCCGTCGTTGCGA
TCGGCGCCGCTGGACTCGCCGGAGCTGATCCGGCTCGCCCCGCGCCCGGCCGCGCAACCGATCCG
ATGCTGTTGGCGTGGACGGTACCGGTGGTGGACCTGGACCCCAACCGCGGCGTTGGCCGCCTTCGAC
CAGCCCGCCCCCGACGTCCGCTACGGCGCGTCCGTCGACTACCTGGCCGAGCTGGCCGTTTTTCGCG
CGCGAGTTGGTCGAGCGTGGTCGCGTGCTGCCCCAGCTGCGCCGCGACACCCACGGCGCGGCCGCC
TGCTGGCGTCCGGTGTTGCAGGGACGCGACGTGGTCGCGATGACCTCGCTGGTCTCGGCGATGCCG
CCGGTCTGCCGCGCCGAAGTTGGTGGGCACGACCCGCACTGGAACCTCGGCTCTGGACGCG
ATGGTCGACGCGCCGCGTGC GCGCGGCGCTGTCACCGATGGACCTGCTGCCCCGCGACGGGGTCGC
TCCAAACGGCATCGGGCCGTGGAGGCTTGGCTGACCGCGTTGACCTGCCCCGACGGCCGGTTCGAC
GCGGAGCCCGACGAAC TCGACGCGCTGGCCGAGGCGTTGCGGCCATGGGACGACGTCGGTATCGGC
ACCGTCGGCCCCGGCGCGGGCGACGTTTTCGGCTGTCCGAAGTCGAGACCGAAAACGAGGAGACGCCC
GCGGGCTCGTTGTGGAGGCTGGAGTTCTTATTGCAGTCGACGCGAGGACCCAGCCTGCTGGTCCCC
GCCGAGCAGGCATGGAACGACGACGGCAGCCTGCGCCGCTGGCTGGACCGGCCGCGAGGAGCTGCTG
CTGACCGAACTGGGCCCCGGCCTCTCGGATTTTCCCCGAGCTCGTCCCGGCGCTGCGCACCGCGTGC
CCGTCCGGGCTTGAGCTCGACGCCGACGGCGCCTACCGATTCTGTGGGTACGGCCGCGGTGCTC
GACGAGGCTGGGTTTGGCGTGCTGCTGCCGTCCTGGTGGGACCGCCGCGCAAGCTGGGCTTGGTC
CTGTCCGCATATACCCCGGTGACGGCGTGGTGGGCAAGGCCAGCAAGTTCGGCCGCGAGCAGCTC
GTCGAGTTCCGCTGGGAGCTGGCCGTGGGCGACGATCCGCTCAGCGAGGAGGAGATCGCGGCGCTG
ACCGAAACCAAGTCCCCGCTGATCCGGCTGCGTGGCCAGTGGGTGGCGCTCGATACCGAACAGCTG

FIGURE 10 (continued)

56/96

CGCCGCGGGCTGGAGTTTTTGGAGCGTAAGCCAACCGGCCGCAAGACCACCGCCGAGATCCTCGCG
CTGGCCGCCAGCCACCCCGACGACGTGGACACCCCGCTCGAGGTCACCGCCGTACGCGCCGACGGC
TGGCTCGGGGACCTGCTCGCCGGGGCCGCCGCGGGCGTCTGCTGCAGCCGTTGGACCCGCCCGACGGA
TTCACCGCGACGCTGCGTCCCTACCAGCAGCGCGGTCTGGCGTGGCTGGCGTTTTTGTCTCTGCTC
GGTTTGGGCAGCTGCCTGGCCGACGACATGGGCCTGGGCAAGACGGTGCAGCTATTGGCCCTGGAA
ACCTTGGAATCCGTTTCAGCGCCACCAGGATCGCGGGCGTTCGGACCCACACTGCTACTGTGCCCGATG
TCGTTGGTGGGCAACTGGCAGCAGGAAGCGGCCAGGTTTGCACCCAACCTGCGGGTGTACGCCAC
CACGGGGGCGCCCGGCTGCACGGCGAGGCGTTGCGCGACACCTCGAGCGCACCGACCTGGTCGTG
AGCACCTATACACCGCCACCCGCGACATCGACGAGCTGTCGGAATACGAATGGAACCGGGTGGTG
CTGGACGAGGCCAGGCGGTGAAGAACAGCCTGTCCCGGGCGGCCAAGGCGGTGCGACGGCTACGC
GCGGCGCACCGGGTTCGCGCTGACCGGGACACCGATGGAGAACCGGCTCGCCGAGCTGTGGTTCGATC
ATGGACTTCCTCAACCCGGGCCTGCTCGGATCCTCCGAACGCTTCCGCACCCGCTACGCGATCCCG
ATCGAGCGGCACGGGCACACCGAACCGGCCGAACGGCTGCGCGCATCGACGCGGCCCTACATCCTG
CGCCGGCTCAAGACCGACCCGGCGATCATCGACGATCTGCCGGAGAAGATCGAGATCAAGCAGTAC
TGCCAACCTCACCACCGAGCAGGCGTCGCTGTATCAGGCCGTCGTCGCCGACATGATGGAAAAGATC
GAAAACACCGAAGGGATCGAGCGGCGCGGCAACGTGCTGGCCGCGATGGCCAAGCTCAAACAGGTG
TGCAACCACCCCGCCCAGCTGCTGCACGATCGCTCCCCGGTTCGGTTCGGCGGTCCGGGAAGGTGATC
CGGCTCGAGGAGATCCTGGAAGAGATCCTGGCCGAGGGCGACCGGGTGTGTGTTTTACCCAGTTC
ACCGAGTTCGCCGAGCTGCTGGTGCCGCACCTGGCCGCACGCTTCGGCCGTGCCGCCCGAGACATT
GCCTACCTGCACGGTGGCACCCCGAGGAAGCGGCGTGACGAGATGGTGGCCCGGTTCCAGTCCGGT
GACGGCCCGCCCATTTTTCTGCTGTCGTTGAAGGCGGGCGGTACCGGGCTGAACCTCACCGCCGCC
AATCATGTTGTGCACCTGGACCGCTGGTGGAACCCGGCGGTTCGAGAACCAGGCGACGGACCGGGCG
TTTCGGATCGGGCAGCGGCGCACGGTGCAGGTCCGCAAGTTCATCTGCACCGGCACCCTCGAGGAG
AAGATCGACGAAATGATCGAGGAGAAAAAGGCGCTGGCCGACTTGGTGGTCAACGACGGCGAAGGC
TGGCTGACCGAACTGTCCACCCGCGATCTGCGCGAGGTGTTTCGCGCTGTCCGAAGGCGCCGTCCGT
GAGTAG

**SEQ ID NO: 50, Mycobacterium bovis BCG Pasteur 1173P2 Myco_SNF2
translated polypeptide**

MLVLHGFWSNSGGMRLWAEDSDLLVKSPSQALRSARPHPFAPADLIAGIHPGKPATAVLLLPSLR
SAPLDSPELIRLAPRPAARTDPMLLAWTVPVVDLDPTAALAAFDQPAPDVRYGASVDYLAELAVFA
RELVERGRVLPQLRRDTHGAAACWRPVLQGRDVVAMTSLVSAMPFVCRAEVGGHDPHELATSALDA
MVDAAVRAALSPMDLLPPRRGRSKRHRAVEAWLTALTCPDGRFDAEPDELDAEALRPWDDVGIG
TVGPARATFRLSEVETENEETPAGSLWRLEFLLQSTQDPSLLVPAEQAWNDDGSLRRWLDRPQELL
LTELGRASRIFPELVPALRTACPSGLELDADGAYRFLSGTAAVLDEAGFGVLLPSWWDRRKLGLV
LSAYTPVDGVVGKASKFGREQLVEFRWELAVGDDPLSEEEIAALTETKSPLIRLRGQWVALDTEQL
RRGLEFLERKPTGRKTTAEILALAASHPDDVDTPLEVTAVRADGWLGDLLAGAAAASLQPLDPPDG
FTATLRPYQQRGLAWLAFLLSSLGLGSLCLADDMGLGKTVQLLALETLESVQRHQDRGVGPTLLLCPM
SLVGNWQQEAARFAPNLRVYAHHG GARLHGEALRDHLERTDLVVSTYTTATRDIDELSEYEWNRVV
LDEAQAVKNSLSRAAKAVRRLRAAHRVALTGTMPENRLAELWSIMDFLNPGLLGSSERFRTRYAIP
IERHGHTEP AERLRASRPYILRRLKTDPAIIDDLEKIEIKQYCOLTTEQASLYQAVVADMMEKI
ENTEGIERRG NVLAAMAKLKQVCNHPAQLLHDS PVGRRSGKVIRLEEILEEILAEGDRVLCFTQF
TEFAELLVPHLAARFGRAARDIAYLHGGT PRKRRDEMVARFQSGDGPPIFLLSLKAGGTGLNLTA
NHVVHLDRWWNPAVENQATDRAFRIGQRRTVQVRKFICTGTLEEKIDEMIEEKKALADLVVTDGEG
WLTELSTRDLREVFALSEGAVGE

FIGURE 10 (continued)

57/96

SEQ ID NO: 51, *Mycobacterium tuberculosis* H37Rv Myctu_SNF2 nucleic acid sequence

ATGCTGGTTTTTGCACGGCTTCTGGTCCAACTCCGGCGGGATGCGGCTGTGGGCGGAGGACTCCGAT
CTGCTGGTGAAGAGCCCGAGTCAGGCGCTGCGCTCCGCGCGGCCACACCCGTTTCGCGGCGCCCGCT
GACCTGATCGCCGGCATAACATCCGGGCAAACCCGCAACCGCCGTTTTTGCTGTTGCCGTCGTTGCGA
TCGGCGCCGCTGGACTCGCCGGAGCTGATCCGGCTCGCCCCGCGCCCGGCCGCGGAACCGATCCG
ATGCTGTTGGCGTGGACGGTACCGGTGGTGGACCTGGACCCCAACCGCGGCGTTGGCCGCCTTCGAC
CAGCCCGCCCCCGACGTCCGCTACGGCGCGTCCGTCGACTACCTGGCCGAGCTGGCCGTTTTTCGCG
CGCGAGTTGGTCGAGCGTGGTCGCGTGCTGCCCCAGCTGCGCCGCGACACCCACGGCGCGGCCGCC
TGCTGGCGTCCGGTGTTCAGGGACGCGACGTGGTCGCGATGACCTCGCTGGTCTCGGCGATGCCG
CCGGTCTGCCGCGCCGAAGTTGGTGGGCACGACCCGCGACGAACCTCGGCTCTGGACGCG
ATGGTCGACGCGCCGCGTGCAGCGCGGCGCTGTCACCGATGGACCTGCTGCCCCCGCGACGGGGTCGC
TCCAAACGGCATCGGGCCGTGGAGGCTTGGCTGACCGCGTTGACCTGCCCGGACGGCCGGTTCGAC
GCGGAGCCCGACGAACCTCGACGCGCTGGCCGAGGCGTTGCGGCCATGGGACGACGTCCGTATCGGC
ACCGTCGGCCCCGGCGCGGGCGACGTTTTCGGCTGTCCGAAGTCGAGACCGAAAACGAGGAGACGCCC
GCGGGCTCGTTGTGGAGGCTGGAGTTCTTATTGCAGTCGACGCGAGGACCCAGCCTGCTGGTCCCC
GCCGAGCAGGCATGGAACGACGACGGCAGCCTGCGCCGCTGGCTGGACCGGCCGCGAGGAGCTGCTG
CTGACCGAACTGGGCCGGGCCTCTCGGATTTTCCCCGAGCTCGTCCCGGCGCTGCGCACCGCGTGC
CCGTCCGGGCTTGAGCTCGACGCCGACGGCGCCTACCGATTCTGTCCGGTACGGCCGCGGTGCTC
GACGAGGCTGGGTTTGGCGTGCTGCTGCCGTCTTGGTGGGACCGCCGCGCAAGCTGGGCTTGGTC
CTGTCCGCATATACCCCGGTTCGACGGCGTGGTGGGCAAGGCCAGCAAGTTCGGCCGCGAGCAGCTC
GTCGAGTTCCGCTGGGAGCTGGCCGTGGGCGACGATCCGCTCAGCGAGGAGGAGATCGCGGCGCTG
ACCGAAACCAAGTCCCCGCTGATCCGGCTGCGTGGCCAGTGGGTGCGGCTCGATACCGAACAGATG
CGCCGCGGGCTGGAGTTTTTGGAGCGTAAGCCAACCGGCCGCAAGACCACCGCCGAGATCCTCGCG
CTGGCCGCGCAGCCACCCCGACGACGTGGACACCCCGCTCGAGGTACCGCCGTACGCGCCGACGGC
TGGCTCGGGGACCTGCTCGCCGGGGCCGCGCGGGCGTTCGCTGCAGCCGTTGGACCCGCCCCGACGGA
TTCACCGCGACGCTGCGTCCCTACCAGCAGCGCGGTCTGGCGTGGCTGGCGTTTTTGTCTCTCGCTC
GGTTTGGGCAGCTGCCTGGCCGACGACATGGGCCTGGGCAAGACGGTGCAGCTATTGGCCCTGGAA
ACCTTGGAATCCGTTTCAGCGCCACCAGGATCGCGGCGTCCGACCCACACTGCTACTGTGCCCGATG
TCGTTGGTGGGCAACTGGCCGCGAGGAAGCGGCCAGGTTTGCACCCAACCTGCGGGTGTACGCCAC
CACGGGGGCGCCCGGCTGCACGGCGAGGCGTTGCGCGACACCTCGAGCGCACCGACCTGGTCGTG
AGCACCTATAACACCGCCACCCGCGACATCGACGAGCTGGCGGAATACGAATGGAACCGGGTGGTG
CTGGACGAGGCCAGGCGGTGAAGAACAGCCTGTCCCGGGCGGCCAAGGCGGTGCGACGGCTACGC
GCGGCGCACCGGGTTCGCGCTGACCGGGACACCGATGGAGAACCGGCTCGCCGAGCTGTGGTTCGATC
ATGGACTTCCTCAACCCGGGCCTGCTCGGATCCTCCGAACGCTTCCGCACCCGCTACGCGATCCCG
ATCGAGCGGCACGGGCACACCGAACCGGCCGAACGGCTGCGCGCATCGACGCGGCCCTACATCCTG
CGCCGGCTCAAGACCGACCCGGCGATCATCGACGATCTGCCGGAGAAGATCGAGATCAAGCAGTAC
TGCCAACTCACCACCGAGCAGGCGTCGCTGTATCAGGCCGTCGTCGCCGACATGATGGAAAAGATC
GAAAACACCGAAGGGATCGAGCGGCGCGGCAACGTGCTGGCCGCGATGGCCAAGCTCAAACAGGTG
TGCAACCACCCCGCCCAGCTGCTGCACGATCGCTCCCCGGTCGGTCGGCGGTCCGGGAAGGTGATC
CGGCTCGAGGAGATCCTGGAAGAGATCCTGGCCGAGGGCGACCGGGTGTGTGTTTTACCCAGTTC
ACCGAGTTCGCCGAGCTGCTGGTGCCGACCTGGCCGCGACGCTTCGGCCGTGCCGCCCGAGACATT
GCCTACCTGCACGGTGGCACCCCGAGGAAGCGGCGTGACGAGATGGTGGCCCGGTTCCAGTCCGGT
GACGGCCCGCCCATTTTTCTGCTGTCGTTGAAGGCGGGCGGTACCGGGCTGAACCTCACCGCCGCC
AATCATGTTGTGCACCTGGACCGCTGGTGGAACCCGGCGGTCGAGAACCAGGCGACGGACCGGGCG
TTTCGGATCGGGCAGCGGCGCACGGTGCAGGTCCGCAAGTTCATCTGCACCGGCACCCTCGAGGAG
AAGATCGACGAAATGATCGAGGAGAAAAAGGCGCTGGCCGACTTGGTGGTCAACGACGGCGAAGGC
TGGCTGACCGAACTGTCCACCCGCGATCTGCGCGAGGTGTTTCGCGCTGTCCGAAGGCGCCGTCCGT
GAGTAG

FIGURE 10 (continued)

58/96

SEQ ID NO: 52, *Mycobacterium tuberculosis* H37Rv Myctu_SNF2 translated polypeptide

MLVLHGFWSNSGGMRLWAEDSDLLVKSPSQALRSARPHFPAAPADLIAGIHPGKPATAVLLLP
SLR
SAPLDSPELIRLAPRPAARTDPMLLAWTVPVVDLDPTAALAAFDQPAPDVRYGASVDYLAELAVFA
RELVERGRVLPQLRRDTHGAAACWRPVLQGRDVVAMTSLVSAMP
PVCRAEVGGHDPHELATSALDA
MVDAAVRAALS
PMDLLPPRRGRSKRHRAVEAWLTALTCPDGRFDAEPDELDALAEALRPWDDVGIG
TVGPARATFRLSEVETENEETPAGSLWRLEFLLQSTQDPSLLVPAEQAWNDDGSLRRWLDRPQELL
LTELGRASRIFPELVPALRTACPSGLELDADGAYRFLSGTAAVLDEAGFGVLLPSWWDRRRKLGLV
LSAYTPVDGVVGKASKFGREQLVEFRWELAVGDDPLSEEEIAALTETKSPLIRLRGQWVALDTEQM
RRGLEFLERKPTGRKTTAEILALAASHPDDVDTPLEVTAVRADGWLGDLLAGAAAASLQPLDPPDG
FTATLRPYQQRGLAWLAFLLSSLGLGSC
LADDMGLGKTVQLLALETLESVQRHQDRGVGPTLLLCPM
SLVGNWPQEAARFAPNLRVYAHHG
GARLHGEALRDHLERTDLVVSTYTTATRDIDELAEYEWNRVV
LDEAQAVKNSLSRAAKAVRRLRAAHRVALTGT
PMENRLAELWSIMDFLNPGLLGSSERFRTRYAIP
IERHGHT
EPAERLRASTRPYILRRLKTDPAI
IDDLPEKIEIKQYCQLTTEQASLYQAVVADMMEKI
ENTEGIERRGNVLAAMAKLKQVCNHPAQLLH
DRSPVGRRS
GKVIRLEEILEEILAEGDRVLCFTQF
TEFAELLVPHLAARFGRAARDIAYLHGGT
PRKRRDEMVARFQSGDGPPIFLLSLKAGGTGLNL
TAA
NHVVHLDRWWNPAVENQATDRAFRIGQRRTVQVRKFI
CTGTLEEKIDEMIEKKALADLVVTDGEG
WLTELSTRDLREVFALSEGAVGE

SEQ ID NO: 53, *Myxococcus xanthus* DK 1622 Myxxa_DK_SNF2 nucleic acid sequence

GTGCGAGCCTGGAGGGGCGTCCTCCGCTGGGCTGCCGCTGGCCTCTCCCTGTCCGCGGCTCGGAGT
CCGACCGGCCACCTCCCAGTGTTTTTCAGGTTTTTCCGTGGCGACCGATGGCGTCGGGCTGTTCGCG
GGTCTGTCTGTTCGGGCCCTTGTCCATCAAGGGCCTGGAGGAGGACCGCTACGAGCGCCTCACGGA
CAACCCGGCAGGCCTGCGGCTCACGGAGCCGGCAATCCCGTGCAGGGGCGCTCGCAGGCCTGCTTG
CGTGTGCCGCTTGCCCCGGACGGAGTTTACATTTCGCAGCGATGCCCCCTCGTGTTCCTGCCCGACGCC
GAGACGCTGTTCTCTGGGGGCCCCGACCGGCTGCCACGTGAGCTCGCCGGCCTGCCGGAGACGGGG
GACCGCGCCTCCGCGCTGCTCGTGACGCCCCGAGGGATTGCGTGAATGCGAGGGGGCACGGGCTGCCC
CTGGCCGCCACCGTCGAGCGGCTCGCGGTGGTGCAAACCTCCGAGGCCGAGTCCTTTCTGGCTCC
ATCGCCCTGTGGACGCTGGCCAGCAAGCTCGCGCTGGAGTTGGTGGCGCGCGAGCGCGTGGTGCCC
ACGCTCCTGCGGCGGGGCGAGCGCATCGAGGCTCGCTGGGCGGGCGGCCCTCTCCGCCACCGAGGAC
GCCGGCCGCGTCGCCGCGCTCGCCCGGAGCATGCCGCCCGGCGCGCACGCCGTCCCCGCAGGCGCC
AGGCCAGGCCGCGCCGTCTGGGCCCCGGACGCCTTGCTGCGCGCCTTCCTCGACGCCACCGTCGAC
GCCTTCGTGCGCGCCGCGCGCGGTGCGCCTTCGTTGCCGGCCCCGGCGCGCGGCCTCGTGGGACGAG
CGCTGGCGCGAGGCGCTCACCGGCGCGCGACGCGACTTCGCGCCGGAGGGCTTCGCCGAGCGCTCC
GTCGTGATGAGCTGACGCGCTGGAGCGAACCCGCGCTCGGCGCCCCGGGACAAGCTGCGCGCCTGC
TTCCGGCTGGAGCCCCCGACGGAGGAGCGCGAGCCCTTCGTGCTGAGCTTCCACCTCCAGTCCCCG
GACGACCCAAGCCTGCTCGTCCCGGCCGCGGACGTCTGGAAGACGCGCGGGGCGCAGCCTGGAGAAG
CTCGGCCGCGCCTTCGCTGACCCGCGAGGAGTCCCTGCTCGAGGCACTCGGCCGCGCCGCCCCGGCTC
TTCCCCCGCTGGCGCTCGTGCTGGAGAGCCCACGTCCCCAGGCGCTCCTGCTCGAGCCCGACACC
GCGTGGACGTTCTCTCGGAGGGCGCCCGCGTGCTCTCAGACGCCGGCTTCGGCGTCATCGTCCCT
GGCGAGCTCACCACCTCGGGCCGACGCCGCCTGCGCCTGCGCATGCGCGTGGGCGCGAGCACGAAG
GCCGCGGGGGCCGTGGTGGCACCGCGGGGCTCGGGCTCGACGCGCTGCTGCGCGTGGACTGGGAC
GCCGTGCTGGGCGACCAACCCCTCTCCGCCAGGAGCTGGCGCTGCTGGCCCAGCGCAAGGCCCCG
CTCGTGCGATTCCGCGGCGAGTGGGTGCGGGTGGATCCCCTCGAACTCGACGCCATCCAGCGCCAC
CTCGCCCAGGGCCCCGGCCGCATGGCGCTGAGCGAGGCGGTGCGGGTGTCCCTGCTAGGCGAAACG
CGCCACGGACAGCTCCCCGTACCGTTCTCGCCACCGGGGCGCTGGAGGAGCGCCTGCGCCTGCTT
CGGGAGGGCGGGGCCACCGCTCAGGACGCCCCCGCGCGCTGCGCGCCACGCTGCGGGCCCTACCAG

FIGURE 10 (continued)

59/96

TCGCGCGGTCTGCACTGGCTGGACACGCTGGCCTCATTGGGGCTCGGCGCCTGCCTCGCGGACGAC
ATGGGCCTGGGCAAGACGGTGCAGGTGCTGGCCTTCCTGCTGCGGCGGCTCGAGCAGGCGCCTGAC
GAGGCGCGCCCCACGCTGCTGGTGGCCCCCACCTCCGTGGTGGGCAACTGGGAGCGTGAGCTCGCC
CGCTTCGCCCCCACCTTGCGCCTGACGCGGCACTACGGCGCCGAGCGCGCCCCGCGCGGGCGAACCGC
TCCCCCGCGCGCCCCGGCGCCGTCTGTGCTCACCACCTACGGCTTGCTGCGCCGGGACGCCGCGCTG
CTCGCGCGCGTGGACTGGGGCGCGGTGGTGTCTGACGAGGCGCAGAACATCAAGAACGCGGCGTCTG
GCTACCGCCCCGCGCGGGCCCCGGGCGTTGCGCGCCAGCCAGCGCTTCGCGCTCACGGGCACGCCGGTG
GAGAACCGCCTGGCGGAGCTGTGGTCCATCCTCGAGTTCGCCAACCCGGGCTGCTCGGGCCGCTG
GAGACGTTCCGGCGGGAGCTGGCGCTGCCCATTTGAACGCCATGGCAATCAGGAGGCCTCGGCCCGG
CTGCGCCGGCTCGTGAGCCCCCTTCGTCTGCGCCGCCTCAAGAGCGACCCGACCATCATCACGGAC
CTGCCCCGCAAGAATGAGATGAAGGTCGTCTGCACGCTCACGCGCGAGCAGGCCTCGCTCTACAAG
GCGGTGGTGGACGAGGAGCTGCGGCGCATCGAGGAGGCCGACGGCATGGAGCGCCGGGGCCGCGTG
CTCGCGCTGCTGCTGTACACGAAGCAGATCGCCAACCAACCCGGCGCAGTACCTCGGGGAGTCCGGG
CCCCTGCCGGGGCGCTCGGGGAAGCTGGCGCGCGTGGTGGAGATGCTCGAGGAGTCCCTGGCCGCT
GGCGACAAGGCGCTCGTCTTCACGCAGTTCCGGGAGATGGGCGACAAGCTGGTGGCGCACCTGTCTG
GAGTACCTGGGCCACGAGGTGCTCTTCCTCCACGGCGGCACGCCCCGCAAGGCGCGCGACGAGATG
GTGCGGCGCTTCCAGGAGGACGTCCACGGTCCGCGTGTGTTCTGCTGTCCGTCAAGGCGGGAGGC
ACGGGGCTCAACCTGACGGCGGCGAGCCATGTGTTCCATTACGACCGCTGGTGGAACCCGGCCGTC
GAGGACCAGGCCACCGACCGCGGTACCGCATCGGGCAGACGCGCGCGGTGCAGGTCCACAAGCTG
GTGTGTGCGGGCACTGTCGAGGAGAAGGTGGACCGGCTGCTCGAACAGAAGCGCCAGCTCGCCGAG
AAGGTCGTGGGCGCGGGCGAGCACTGGGTGACCGAGCTGGACACGACGGCGCTGCGCGAGCTGTTT
TCGCTGTCCGAGGGCGCCGTGGCGGACGATGGCGACGCGGAAGGGGAAGACGACGCGCGGGTGC
GCCCCGCGACGGCGCGGCCGTGCGAGCGCGAAGGCGGTGTCGCGATGA

**SEQ ID NO: 54, *Myxococcus xanthus* DK 1622 Myxxa_DK1622_SNF2
translated polypeptide**

VRAWRGVLRWAAAGLSLSAARSP TGHLPVFSGFSVATDGVGLFAGLSVRALVHQGPGGGPLRAPHG
QPGRPAAHGAGNPVQGRSQACLRVPLARTEFTFAAMPLVFLPDAETLFLWGPDRLPRELAPETG
DRASALLVTPEGLRECEGHGLPLAATVERLAVVQTSEAESFPGSIALWTLASKLALELVARERVVP
TLLRRGERIEARWAAALSATEDAGRVAALARSMPGAHAVPAGARPGRAVWAPDALLRAFLDATVD
AFVRAARGAPSLPARRAASWDERWREALTGARRDFAPEGFAERSVVDLTRWSEPALGARDKLRAC
FRLEPPTEEREPEFVLSFHLQSPDDPSLLVPAADVWKTRGRSLEKLGRAFRDPQESLLEALGRAARL
FPPLALVLESPPRQALLLEPDTAWTFLSEGARVLS DAGFGVIVPGELTTSGRRLRLRMRVGASTK
AAGAVGGTAGLGLDALLRVDWDAVLGDQPLSAQELALLAQRKAPLVRFRGEWVAVDPLELDAIQRH
LAQGPGRMALSEAVRVSLGETRHGQLPVTVLATGALEERLRLLLREGGATAQDAPRALRATLRPYQ
SRGLHWLDTLASLGLGACLADDMGLGKTVQVLAFLRLRLLEQAPDEARPTLLVAPTSVVGNNWERELA
RFAPTLRLTRHYGAERARAANRFPRAPGAVVLTITYGLLRRDAALLARVDWGAVVLDEAQNINKNAAS
ATARAARALRASQRFALTGTPVENRLAELWSILEFANPGLLGPLETFRRELALPIERHGNQEASAR
LRRLVSPFVLRRLKSDPTIITDLPKKNEMKVCTLTREQASLYKAVVDEELRRIEADGMERRGRV
LALLLYTKQIANHPAQYLGESGPLPGRSGKLARVVEMLEESLAAGDKALVFTQFREM GDKLV A H L S
EYLGHEVLFLHGGT PRKARDEMVR RFQEDVHGPRV FVLSVKAGGTGLNLTAASHVFHYDRWWNP AV
EDQATDRAYRIGQTRAVQVHKLVCAGTVEEKVDRLLLEQKRQLAEKVVGAGEHWVTELDTTALRELF
SLSEGAVADDGDAEGEDDARVRAPRRRGRASAKAVSR

FIGURE 10 (continued)

60/96

SEQ ID NO: 55, *Nocardia farcinica* IFM 10152 Nocfa_IFM\10152_SNF2
nucleic acid sequence

ATGGGTGGGCGCCGGCGGCCCGGGGTGTCGGGTGCCACCTGCTTGGATGGACGGATGCTGCACGGA
CTGTGGTCGCCGGGTTCGGGCCTGGTGCTGTGGACCGAGGGCGAGGTGCCGCCCGCGCTGCCCGAC
CCGGCCGGTGCGTTGCTGCGCGCATCGCGGTTCCGGCATCGGGCGCAGGTGCTGGTGCCGGGGCCCC
GCCGGCCACAGCTCACGCAGGTGCGCGCGCACGCCCTGGTGCCACAGGCCGCGGTTCGACGTGCTG
CGGCAGCGGTTACCCGTCGAATCGGTGGCGGGTGACCTGCGCTTTCTCGCTCACGTGCCCGACGGG
ATCGATCGGTGGGTGCGGGCCGGTCGCGTGGTGCCCGACCTGCACCGGGCCGACGGACAGTGGTGG
GCGCGCTGGCGGCTGGTCGGCGGTGCCCGGCAGCGGGCCTGGCTGGCCGAACCTCGCGGTGGCGATG
CCCGCGGCGCTGCGGGTGGCCGGGCAGCCCGCGGCGGTGCTCGACGATCTGGTCACCGAGCTGACC
GATCCGATCGTGCGCACCCAGGCTCGCCGACGCGCCGGTGACGCAACCCGCTGGTGCGCGCACTGGTG
CGGGACCAGCCGCTCGAGACGGGTAGCCACCAGCTGGCCGAGGTGCTGCGGCGCTGGCGCGAGAGC
CTCACCGTCGACGAGCCGAGCTGGTGTTGCGGCTGCTGGAACCGGACGGGGAGACCGGTATCGAC
GGGGACGGCGGGGACGACCGGGACGACACCGTGGCGCTGTGGCGGCTGGAGGTCTGCCTCCGCACC
GAGGGCGAGGCCCGGCCCGGTGCCGGCGACCGCCGACCCGAACCTGCTGCGCATCGCCGTTCGAG
CAGCTCGGCCGGGCGCAGCGGGCCTACCCCGGCTGCGCGATCTGCCCGGCGATCCGCACAGCCTC
GACCTGCTGTTGCCACCGAGGTGGTGCCGATCTCGTCGCGCACGGTGCGCAGGCGTTGCGCGAG
GCGGGGGTGCGGCTGCTGCTGCCGCGCGCCTGGACCATCGCCGAACCCACCCCTGCGGCTCGCGGTG
AGCAGCGCCGCGCCCCGCCGCGGAGAGCACCGTGGGCATGCAGGGTCTGCTGTCTATCGGTGGGAA
CTGGCGGTCCGGCGACAAGGTGCTCACCCGCGCCGAGATGGAGCGCCTGGTCCGCGCCAAATCCGAC
CTGGTGCAAGTTGCGCGGGGAATGGGTGCAGGCCGACCACAAGGTGCTCGCCGCCGCCGCCCGCTAC
GTCGCCGCGCATCTGGACACGTCCCGGTACCCCTCGCCGACCTGCTCGGCGAGATCGCCGCCACC
CGCGTCGACAAGGTGCCGCTACCGAGGTACCGCCACCGGCTGGGCGGGCGAGTTGTTTCGACGGC
GGCCGCGAGCCGGTGGCGACCCCGGGTGGGCTGAAGGCGCAGCTGCGCCCGTATCAGCTGCGCGGC
CTGAGCTGGCTGGCGACGATGAGCCGGATGGGCTGCGGCGGCATCCTCGCCGACGACATGGGTCTC
GGCAAGACGGTGCAGGTGCTGGCCCTGCTGGTGCACGAGCGCGAGACCAGCACGGCACCGCCCGGC
CCGACACTGCTGGTGTGCCCGATGTCGGTGGTTCGGCAACTGGCAGCGCGAGGCGCAGCGGTTCGCC
CCCGGGCTGCGGGTGGTGGTGCACCACGGCGCCGACCGCCGTCGCGACGCCGAACCTCGATGCCGCG
GTGGCGGATTTCGGACCTGGTGCTCACACCTACGCCATCCTGGCCAGGGATGCGGCCGAACCTGTCG
CGCCAGTCGTGGGACCGGGTGGTGCTCGACGAGGCGCAGCACATCAAGAACGCCGCGACCAAGGCAG
GCACGTGCCGCCCGTGCCCTGCCGGCCCGGCATCGCCTGGCGCTCACCGGAACCCCGGTGGAGAAC
CGGCTCGAAGAGTTGCGCTCGATCATGGATTTTCGCGGTGCCCAAGCTGCTCGGTACCGCACCGACC
TTCCGCGCCCCGTTTCGCCGTCCCCATCGAACGCGGGCAGGATCCCAACGCCCTGTCCCGCCTGCGC
TTCCTCACCCAACCGTTTCGTGCTGCGCCGGGTCAAGGCCGATCCGGCGGTTCATCGGCGATCTGCCC
GACAAGCTCGAGATGACGGTGCGGGCGAACCTGACCGTCGAGCAGGCCGCCCTGTACCAAGCCGTC
GTCGACGACATGCTGGTGAAACTGCGCAGTGCCAAGGGCATGGCCCGCAAGGGTGCGGTGCTCGGC
GCGCTCACCCGGCTCAAGCAGGTGTGCAACCATCCCGCGCACTTCCTCGGTGACGGTTCCCCGGTG
CTGCATCGCGGCAGGCACCGCTCCGGCAAGCTCGCCTTGGTTCGAGGACGTGCTCGACACCGTCGTC
GCGGACGGGGAGAAGGCGTTGCTGTTACCCAGTTCCGTGAGTTTCGGCGACCTGCTCGCGCCCTAT
CTGTCCGAGCGGTTTCGGCGCGCCGATCCCGTTCCTGCACGGCGGCGTGACCAAGAAGAACCGGGAC
ACGATGGTCGAGCGCTTCCAGTCCGGCGACGGCCCGCCGGTTCATGCTGCTGTCCCTCAAGGCCGGC
GGCACCGGGCTCACCCCTCACCGCCGCCAATCACGTGGTGCACCTGGATCGCTGGTGGAATCCGGCG
GTGGAGAACCAGGCCACCGATCGCGCCTTCCGCATCGGCCAGCGCCGCGACGTCCAGGTGCGCAAG
CTGGTCTGCGTCGACACCATCGAGGAACGGATCGACGAGATGATCACCGGCAAGAGCAGGCTCGCG
GACCTGGCCGTGGACGCGGGGGAGAACTGGATCACCGAGCTGGGCACCGAGGAGCTGCGCGAGTTG
TTCACCCCTCGGCGCCGAGGCGGTGGGGGAGTGA

FIGURE 10 (continued)

61/96

SEQ ID NO: 56, *Nocardia farcinica* IFM 10152 Nocfa_IFM_10152_SNF2 translated polypeptide

MVGAGGPPGVGATCLDGRMLHGLWSPGSGLVLWTEGEVPPALPDPAGALLRASRFRHRAQVLVPGP
AGPQLTQVRAHALVPQAAVDVLRQRLPVESVAGDLRFLAHVADGIDRWVRAGRVPDLHRADGQWW
ARWRLVGGARQRAWLAELAVAMPAALRVAGQPAAVLDDLVTELTDPPIVRLADAPVTHPLVRALV
RDQPLETGSHQLAEVLRRWRESLTVDEPELVLRLLLEPDGETGIDGDDGDDRDDTVALWRLEVCLRT
EGEAPAPVPATADPNLLRIAVEQLGRAQRAYPRLRDLPGDPHSLDLLLPTEVVADLVAHGAQALRE
AGVRLLLPRAWTIAEPTLRLAVSSAAPAAESTVGMQGLLSYRWELAVGDKVLTRAEMERLVRAKSD
LVQLRGEWVQADHKVLA AAAARYVAAHLDTS PVT LADLLGEIAATRVDKVPLTEVTATGWAGELFDG
GREPVATPGGLKAQLRPYQLRGLSWLATMSRMGCGGILADDMGLGKTVQVLALLVHERETSTAPPG
PTLLVCPMSVVGNWQREARFAPGLRVLVHHGADRRRDAELDAVADSDLVLTYYAILARDAEELS
RQSWDRVVLDEAQHIKNAATRQARAARALPARHRLALTGTPVENRLEELRSIMDFAVPKLLGTAPT
FRARFAVPIERGQDPNALSRLRFLTQPFVLRRVKADPAVIGDLPDKLEMTVRANLTVEQAALYQAV
VDDMLVKLRS AKGMARKGAVLGALTRLKQVCNHPAHFLGDGSPVLHRGRHRSGLKALVEDVLDTVV
ADGEKALLFTQFREFGDLLAPYLSERFGAPIPFLHGGVTKKNRDTMVERFQSGDGPPVMLLSLKAG
GTGLTLTAANHVVHLDRWWNPAVENQATDRAFRIGQRRDVQVRKLVCVDTIEERIDEMITGKSRLA
DLAVDAGENWITELGTEELRELFTLGAEAVGE

SEQ ID NO: 57, *Nodularia spumigena* Nodsp_SNF2 nucleic acid sequence

ATGGCAATTTTACACGGTAATTGGTTAGTAAGAAATCAAATGGTTGTTTATTTATTTGGGGTGAA
ACTTGGCGTTCATCACGAGTCGATTTTGTCTCTGAATGTATCTCAAGATATACCACTACATCCATTG
GTAATGTCACCAATTGATTTGAGTGAGTTGTAAAGTTATCATAATATCAAATTCCTAGCTTAATA
CAGCAATCCCAAGTTGCTTTATCTGGCACTGGGCGAACTCGTAAAAGTACAAGTACTACTAAATTT
AGCTGGACAACCTCACTCTCTAATCATTGATTTACCAACTCATATCTCAGAAAATAATCCCCAAGAA
ATAGAATTTATTTCCCCTTTGCATTCTGCTACTTTGGGTTCTGAAATAAATTCACCCCAATATCTC
CAACCGTGGCGAGTCGAGGGTTTTTGTCTCAACCCCACTGAAGCGATAAAATTTCTCGCTGCTGTT
CCTTTAAATGCTGCTAGAGAAGAAGATACTTTGTTTCGGTGGAGATTTACGTTTTTTGGTCACAAATT
GCCCCGTTGGAGTTTGGATTTAATCTCTCGGTGTAAGTTTTTGCCAACTATTCAAAGACAGTTTGAT
AGTTCTATTGTTGCTAGGTGGCAAGTGCTTTTAGACAGTGCAATAGATGGAACACGCCTGGAAAAA
TTTTCTGCAAAAATGCCATTAGCTTGTCGTACTTATCGGAAGGGAATGGGGAGTGGGGAGTGGGGA
GTGGGGAGTGGGGAGGAATCTTCCCCATCCATAATGTATGTAGATTTTCCAACCTGAACCCCAAGGAA
CTATTATTAGGATTTCTCAACAGTACCATAGATGCCCAAGTGCGAGAAATGTTAGCTTCTCAACCT
CTACTAGAACTAGAGTGATGGCATCTTTACCATCTGCGGTGCGACAGTGGTTGCAAGGTTTAACC
AGTGCATCTCACACAGTGAATGCAGATGCAATGGAAGTAGAAAGATTAGAAGCAGCCCTGAAATCT
TGGACTATGCCGTTGCAATATCAACTGGTAGGAAAACCCTCGTTTTCGCGCCTGTTTTCAACTGCTT
CCCCCTGCTTCTGGGGCAACAGATTGGATATTGGCATATTTTCTCCAAGCTGCGGATGATGAAAAT
TTATTAGTGGATGCGGCAACTATTTGGCATCACCCAGTTGAACAATTAGTTTATCAAATCGCACC
ATTGATCAACCCCAAGAACTTTATTGCGGGGCTTGGGTTTAGCTTCGCGATTATATCCAGTTCTT
ACACCGAGTTTAGAAACAGAATATCCCCAATGTTGTGCGCTCAACCCATTACAAGCTTATGAATTT
ATCAAGTCTGTAGCTTGGCGATTTGAAGATAGTGGTTTGGGGGTAAATTTTACCTCCTAGTTTGACT
AACCGCGAAGGATGGGCGAACC GTTTGGGGTTAAAAATTAGTGCTGAACTCAAAAGAAAAAACAG
GGACGCTTGGGTTTACAAAGTTTACTGAATTTTCAATGGCAATTGGCAATTGGTGGACAAACAATT
TCTAAAACCGAGTTTAATAAACTGGTAGCTTTAAATAGCCCACTGGTAGAAATTAACGGCGAATGG
GTGGAATTGCGACCCCAAGGATATTAAAACAGCACAGACATTTTTTGTCTTCTCGTAAAGACGAAATG
ACGCTTTCTTTGGAAGATGCTTTACGCCTCAGTTCTGGCGATACCCAAGCGATTGAAAAGTTACCT
GTGGTCAGTTTTGAAGCATCTGGGACATTGCAAGAGTTAATTGGGGCGTTAACCAATAATCAAGCC
ATTTCAACCCCTCCCAACACCTGCAAATTTTCAAGGACAGTTACGACCTTATCAAGAAAGAGGGGCG

FIGURE 10 (continued)

62/96

GCTTGGCTGGCTTTCTTAGAACGTTGGGGTTTAGGTGCTTGTTTGGCTGATGATATGGGGCTGGGA
AAAACAATTTCAGTTAATTGCCTTTTTTACTGCACCTCAAAGAACAAGACGCACTGGAAAATCCCACA
TTACTTGTTTGTCCGACTTCTATTTTAGGTAAGTGGGAACGGGAAATTAAAAAATTTGCTCCTACT
CTCAAAGTTTTACAGCACCCACGGCGATAAACGTCTCAAAGGTAAAGCGTTTGTAGAAGCAGTCAAA
AAACACGATGTAATTATTACCAGTTACTCACTCGTTCACCGGGATATTAAATCTTTGCAGAGTGTC
GATTGGCAAACAGTTGTATTAGATGAAGCCCAGAATGTGAAAAATCCTGAAGCTAAACAATCGCAG
GCTGTGAGGGGATTAAAACTACATTTTCGCATAGCTTTAACAGGGACACCAGTAGAAAAACAACTG
CAAGAATTGTGGTCTATTTTAGATTTTCTTAATCCTGGGTATTTGGGAAATCGTCAATTTTTCCAG
AGACGGTTTGTCTATGCCAATTGAAAAGTATGGTGATACAGCATCTTTAAATCAATTGCGGGGTTTA
GTTCAACCGTTTATTCTACGTCGTCTGAAAACAGATCGTGATATTATTCAAGATTTGCCAGAAAAG
CAAGAAATGACGGTTTTTTTGTGGGCTTGCGGCTGAACAAGCTGCACTTTATCAACAAGTAGTTGAA
GCATCTTTAGTAGAAATTGAATCTGCTGAGGGTTTGCAACGTCGAGGGATGATTTTAGCTTTACTT
GTGAAACTTAAACAAATCTGTAATCATCCAGCCCAATATTTGAAAGCCGCGACATTACAAGAACAT
AGTTCTGCTAAACTGCAACGGCTAGATGAAATGTTAACGGTAGCTTTGGAGGAAGGAGATAGGGCT
TTAATTTTCACTCAATTTGCTGAATGGGGTAAGTTATTAAAAGCTCATTTACAACAAACACTTGGG
AAAGAAATATTCTTTTTTATATGGTGGTAGCAGTAAAAACAACGCGAGGAAATGATTGACCGTTTC
CAACATGACCCCCAAGGACCTCCGATTATGATTCTTTCTTTAAAGCGGGTGGGGTAGGCTTGAAT
TTAACCAGGGCTAATCATGTATTTCACTTTGATAGATGGTGGAATCCCGCAGTGGAAAATCAAGCG
ACAGATAGAGTATTTTCGTATTGGTCAAACCCGGAATGTGCAAGTGCATAAATTTGTCTGTACTGGC
ACATTAGAAGAAAAAATTCATGACATGATTGAAAGTAAAAACAATTAGCGGAACAAGTAGTTGGT
GCTGGTGAGGAGTGGCTGACTGAAATGAATACTGACCAATTGCGTGATTTACTCATTCTTGATCGC
AGTGCCATAATTGATGAGGATGAAGTTTAA

SEQ ID NO: 58, Nodularia spumigena Nodsp_SNF2 translated polypeptide

MAILHGNWLVRNQNGCLFIWGETWRSSRVDFALNVSQDIPLHPLVMSPIDLSELLSYHNIKIPSLI
QQSQVALSGTGRTRKSTSTTKFSWTHSLIIDLPTHISENNPQEIEFISPLHSATLGSEINSPQYL
QPWRVEGFCLNPTEAIKFLAAVPLNAAREEDTLFGGDLRFWSQIARWSLDLISRCKFLPTIQRQFD
SSIVARWQVLLDSAIDGTRLEKFSKMPPLACRTYRKGMGSGEWGVGSGEESSPSIMYVDFPTEPQE
LLLGLFNSTIDAQVREMLASQPLLETRVMASLPSAVRQWLQGLTSASHTVNADAMEVERLEAALKS
WTMPLQYQLVGKPSFRACFQLLPASGATDWILAYFLQAADDENLLVDAATIWHHPVEQLVYQNRT
IDQPQETLLRGLGLASRLYPVLTPSLETEYPQCCRLNPLQAYEFIKSVAWRFEDSGLGVILPPSLT
NREGWANRLGLKISAETQKKKQGRGLGLQSLNLFQWQLAIGGQTISKTEFNKLVALNSPLVEINGEW
VELRPQDIKTAQTFFASRKDEMTLSLEDALRLSSGDTQAIEKLPVVSFEASGTLQELIGALTNNQA
ISPLPTPANFQGQLRPYQERGAAWLAFLERWGLGACLADDMGLGKTIQLIAFLLHLKEQDALENPT
LLVCPTSILGNWEREIKKFAPTLKVLQHHGDKRLKGKAFVEAVKKHDVITTSYSLVHRDIKSLQSV
DWQTVVLDEAQNVKNPEAKQSQAVRGLKTTFRIALTGTPVENKLQELWSILDFLNPGYLGNRQFFQ
RRFAMPIEKYGDTASLNQLRGLVQPFILRRLKTDRDI IQDLPEKQEMTVFCGLAAEQAALYQQVVE
ASLVEIESAEGLORRGMILALLVKLKQICNHPAQYLKAATLQEHSSAKLQRLDEMLTVALEEGDRA
LIFTQFAEWGKLLKAHLQQTLGKEIFFLYGGSSKKQREEMIDRFQHDPQGPPIMLS LKAGGVGLN
LTRANHVVFHFDRWWNPAVENQATDRVFRIGQTRNVQVHKFVCTGTLEEKIHDMIESKKQLAEQVVG
AGEEWLTEMNTDQLRDLLILDRSAIIDEDEV

FIGURE 10 (continued)

63/96

SEQ ID NO: 59, *Nostoc* sp. PCC7120 Nos_sp_PCC7120_SNF2 nucleic acid sequence

ATGGCAATTCTACACGGTAGTTGGATATTAAATGAGCAGGAGAGTTGTTTTATTTATTTGGGGGGGAA
ACTTGGCGATCGCCACAAGTGGATTTTAAATTTTGCGGAGATATCCCTCAATCCCTTGGCGCTGTCT
GCACTGGAATTAAGTGAGTGGTTGCAGTCTCAACATCAGGCGATCGCTAAGTTGTTACCGCAACAA
TTGGAAAAACGAACCTCCAAAGCAGCAAGTTCTGTAAAAATAAATTTATTAACTCATTCACAAATA
ATTGCCCTGCCAACGGAAATTTCCCAACCTCGTAAAAAAGAAACCATTTTAAATTTCTCCTGTGCAT
TCTGCCGCTTTAGCATCTGAGTCAGACTCTGAAGTTTATTTACAACTTGGCGTGTAGAAGGTTTT
TGTCTTCCTCCTAGTGCAGCAATTAAATTGCTAACTTCTTTACCTTTAAATATAACTAGTGGGGAG
AATGCTTTTTTTAGGTGGAGATTTACGTTTCTGGTCACAAATTGCCCGTTGGAGTTTAGATTTAATT
TCTAGGTCTAAGTTTCTCCCAATTATCCAACGACAACCTAATAATTCTGTAAAGTGCTAAATGGCAA
GTACTTTTAGATAGTGCCGTAGATGGAACCTCGTTTAGAAAAGTTTGCTGCGAAGATGCCCTTGGTT
TGTCGGACTTATCAAGAAATTGGGAGTGGGGAATCTCCTATATATATAGATTTTCCTAGTCAGCCG
CAGGATTTAATCTTGGGTTTTCTCAATAGTGCGATAGATACGCAATTGCGGGAGATGGTGGGGAAT
CAGCCTGTGGTGGAACTCGGTTGATGGCATCTTTACCATCGGCGGTGCGACAGTGGTTGCAAGCG
TTAATTGCTGCATCTAATTCAATTGATGCAGATGCTGTTGGTTTAGAAAGGCTGGAAGCGGCGCTC
AAGGCTTGGACGATGCCGCTACAATATCAACTAGCAAGTAAAAATCAATTTTCGCACTTGTTTTGAA
TTACGTTCTCCAGAACCAGACGAACTGAATGGACGCTGGCGTATTTCTGCAAGCAGCCGATGAT
CCAGAATTTTGTAGTAGATGCGGCGACTATTTGGCAAAATCCTGTTGAACAGCTAATTTATCAACAG
CGAACGATTGAAGAACCCCAAGGAAACGTTTTTTCGAGGTTTGGGGTTAGCTTCTCGATTGTATCCG
GTCATTGCCCCCACTTTAGATACAGAATCACCCCAATTTTGTCTCTCAAGCCCATGCAGGCTTAT
GAATTTATCAAGGCTGTGGCTTGGCGATTTGAAGATAGCGGCTTAGGGGTGATTTTACCTCCTAGT
TTGGCGAATCGTGAAGGCTGGGCAAATCGCTTGGGTTTGAAAATCTCCGCCGAAACGCCGAAGAAA
AAACCAGGACGCTTAGGATTGCAGAGTTTGCTCAATTTCCAATGGCACTTAGCGATTGGTGGGCAA
ACTATTTCTAAAGCTGAATTTGACAGACTGGTAGCTTTAAAAAGCCCATTTGGTAGAAATTAACGGC
GAGTGGGTGGAATTACGTCCCCAAGATATCAAAACAGCTGAAGCCTTTTTTACTGCGCGTAAAGAC
CAAATGGCCTTATCTTTAGAAGATGCCTTACGTCTAAGTAGTGGCGATACACAAGTAATTGAGAAA
TTACCAGTAGTCAGCTTTGAAGCCTCTGGCGCATTACAAGAATTGATTGGGGCGCTGACAAATAAT
CAAGCAGTTGCACCATTACCTACGCCGAAAACTTCCAAGGACAGTTACGTCCTTATCAAGAAAGG
GGTGCGGCTTGGTTGGCGTTCTTCGAACGCTGGGGTTTAGGTGCTTGTCTCGCCGACGACATGGGA
CTGGGAAAAACGATACAGTTCATTGCTTTCCTTCTCCATCTTAAAGAACAGGATGTATTAGAAAAA
CCAACTTTACTAGTGTGTCCTACTTCTGTTTTAGGTAACCTGGGAACGAGAGGTGAGAAAATTTGCA
CCTACACTTAAAGTTCTCCAGTATCATGGTGACAAACGTCCTAAAGGTAAAGCATTTTCAAGAAGCA
GTAAAAAAACATGATTTAGTTATTACAAGTTACTCATTAATTCATAGAGATATCAAATCATTGCAG
GGTATTCCTTGGCAAATAATTGTTTTAGATGAAGCCCAAAATGTGAAGAATGCGGAAGCCAAACAA
TCACAAGCAGTCAGACAATTAGAAACAACATTTTCGTATTGCTTTAACAGGTACACCAGTAGAAAAT
AGACTACAAGAACTTTGGTCAATTTTAGATTTTCTTAATCCTGGTTACTTAGGTAATAAGCAATTC
TTTCAAAGACGTTTTTGCTATGCCAATTGAAAAGTATGGTGATGCAGCATCTTTAAATCAATTGCGT
GCTTTAGTGCAACCATTTATTCTGCGTCGGCTGAAAACAGACCGTGATATTATTCAAGACTTGCCC
GATAAGCAAGAAATGACAGTATTTTGTGGTTTGAAGTGGAGAACAAGCTGCACTTTATCAAAAAGCG
GTAGAAACATCTTTAGCAGAAATTGAATCAGCCGAAGGATTGCAACGCCGAGGGATGATTTTAGCT
TTATTAATTAAACTCAAACAAATCTGCAATCATCCAGCCCAATATCTGAAAATAAATACATTAGAA
CAACACAGTTCTGGAAAACCTGCAAAGATTAGAAGAAATGTTAGAAGAGGTGTTAGCAGAGAGTAAT
ACTTACGGTGTTGCCGGTGCGGGACGTGCTTTGATTTTTTACCCAATTTGCAGAATGGGGTAAGTTA
CTCAAACCACATTTAGAAAAACAACCTAGGGCGGGAAATATTTTTCTTATATGGTGGTACGAGTAAA
AAGCAACGAGAAGAAATGATTGACCGTTTTCAACACGACCCCCAAGGGCCACCAATTATGATTCTC
TCCCTCAAAGCAGGTGGTGTAGGGTTGAACTTAACCAGGGCAAATCATGTATTTCACTTTGATAGA

FIGURE 10 (continued)

64/96

TGGTGGGAATCCAGCCGTAGAGAATCAAGCTACAGACCGCGTATTTTCGCATTGGTCAAACCTCGCAAT
GTACAGGTGCATAAATTTGTTTGTAAATGGCACCTTAGAAGAGAGAAAATTCACGACATGATTGAAAGT
AAAAACAACCTAGCGGAACAGGTTGTTGGAGCAGGCGAAGAATGGTTAACTGAATTAGATACAGAT
CAACTCCGCAACTTACTGATACTTGATCGTAGTACAGTAATTGATGAAGAAGCAGATTGA

SEQ ID NO: 60, Nostoc sp. PCC7120 Nos_sp_PCC7120_SNF2 translated polypeptide

MAILHGSWILNEQESCLFIWGETWRSPQVDFNFAEISLNPLALSALELSEWLQSQHQAIAKLLPQQ
LEKRTSKAASSVKINLLTHSQIIALPTEISQPRKKETILISPVHSAALASESDSEVYLQTRVEGF
CLPPSAAIKLLTSLPLNITSGENAFLGGDLRFWSQIARWSLDLISRSKFLPIIQRQPNNSVSAKWQ
VLLDSAVDGRLEKFAAKMPLVCRTYQEIGSGESPIYIDFPSQPQDLILGFLNSAIDTQLREMVGN
QPVVETRLMASLPSAVRQWLQALIAASNSIDADAVGLERLEAALKAWTMPLQYQLASKNQFRTCFE
LRSPEPDETEWTLAYFLQAADDPEFLVDAATIWQNPVEQLIYQQRTIEEPQETFLRGLGLASRLYP
VIAPTLDTESPOFCHLKPMQAYEFIKAVAWRFEDSGLGVILPPSLANREGWANRLGLKISAETPKK
KPGRLGLQSLNLFQWHLAIGGQTISKAEFDRLVALKSPLVEINGEWVELRPQDIKTAEAFFTARKD
QMALSLEDALRLSSGDTQVIEKLPVVSFEASGALQELIGALTNNQAVAPLPTPKNFQGLRPYQER
GAAWLAFLEWRGLGACLADDMGLGKTIQFIAFLHLKEQDVLEKPTLLVCPTSVLGNWEREVRKFA
PTLKVLQYHGDKRPGKAFQEAVKKHDLVITSYSLIHRDIKSLQGIWQIIVLDEAQNVKNAEAKQ
SQAVRQLETTFRIALTGTPVENRLQELWSILDFLNPGYLGNKQFFQRRFAMPIEKYGDAAASLNQLR
ALVQPFILRRLKTDRIIQDLDPKQEMTVFCGLTGEQAALYQKAVETSLAEIESAEGLRGRGMILA
LLIKLKQICNHPAQYLKINTLEQHSSGKLQRLEEMLEEVLAESNTYGVAGAGRALIFTQFAEWGKL
LKPHLEKQLGREIFFLYGGTSKKQREEMIDRFQHDPPQPPIMILSLKAGGVGLNLTRANHVFHFDR
WWNPAVENQATDRVFRIGQTRNVQVHKFVCNGTLEEKIHDMIESKKQLAEQVVGAGEEWLTELDTD
QLRNLLILDRSTVIDEAD

SEQ ID NO: 61, Nostoc sp. PCC7120 Nos_sp_PCC7120_SNF2 II nucleic acid sequence

ATGAAAGTCCTTCATGGCTCGTGGATACCAAACCAATATAGCGATTTTGTGCAGTCTGGAGCATTT
TATCTATGGGTAGAAACTCCGATTAATAACAAAAAGCGTACTCATACACAAGTTCATCCCGGACAT
CTATCTTCTCTTGAATTACTCAATTTTCTGACTCAAACCTTTGGGGATTAAAGAACTGAAGCGCAA
TTAAAACAACGGATATGTTCTAAATATTTTGCCCTACCAACTGCTAATAATGAGCCATTACCTTCA
CCAGAGTTAGTCAAATATTTAGAAGTAGAAGTTCCTGAAGAGTATGAAAATTTTCAATATTGGCAG
GTAACCTTGTTATGAAACTGTTACTTCTGTGAAAGCAGTGATAGCAATTAATATTATTAAATTACTC
AAAGATATTCATTTTTTTAGCCCTGTACAATGCTAGTGAATTTCAATTAGGGTCAGATTTATTATTT
TGGTATCATTATACGCAATCATTTAGACAAATAATTAAGGATCAATATATTCCATCTTTAAAA
TATAGAGCGAACGCAGCGACTACAAAGAAAAAACCTAAACAACCAACCCCAAGGATTTGAAATATAT
GCTGGTTGGGAAATAATTTCCGAGCAATACGAAGCCAATATTCAAAAATATATTGAATATATGCCA
TTGATTTGTGTAGCAGGTAACAGCACACAACTGATAAATTAGAATTTTTTTGCTCCAGAACTCTA
TTACGCCACTTCAGCGAGTATCTGCTTAATAATTTAGTGAGTAAGACACCATTGACCGCAGCATTT
GAAAAACAAATTGATGATTCTTTAATTCATATTGTCTTTATCCCCAAAAACACAACCCACTCAAA
ACCCATACTGCTCTCCAAGAGTATCAGCAGTGGTTGGGATGGAAAAACAGGATTATCCGTACTCAA
GCTGAATCACCATTTCATCTTTGCTTCCAATTACATTCACCTGATGCTGAACAAATTGACAATTGG
CAGATGCAATTTTTTAGTATCAAGTAAAAAAGATCCGTCTCTAAAATTAGCTTTGGCAGATTACTGG
ATAATGAATTCCAAAACCAAAGCTGGTGTACATAAAGAGTTTGGCAAAGATTTTCGATACTAATTTA
CTGCTGAATTTAGGCTATGCAGCAAGAATGTATCCCAAACCTTTGGCAAAGTTTAGAAACGGACTCT
CCCACAGGAATGCAGCTAAGTTTAGATGAGGCGTTTGATTTTCTCAAAGATAGTGCTTGGGTGTTG
GAAGACTCAGGATTTAAGGTCATTGTCCCGGCTTGGTATACTCCGGCTGGTCGTCGTCGTCGCGAAA
ATCCGCCTCAAAGCTTCTAGTGGTCGCAAGGTAGCTGCTACGGTAGGGGAAAGCAAAAGTTATTTTC

FIGURE 10 (continued)

65/96

GGTTTAGATTCACTAGTGCAGTATCAGTATGAATTAGCAATTGGAGAGCAAACCTCTCACACCTCAA
GAATGGGAACAATTGATTAATACTAAAGCACCCTAGTGCATTTTCGCGGTCAATGGATGGAATTA
GACCGGGATAAAATGCAGCAGTTATTAGAATTTTGGCAGTCCCACGGCGATGAACAGCCCCAAATG
AGCTTGTTAGAGTTCATGCAACGCAGCGCCCAAGGGGAAGATGACTGGGAAATTGAATATGATGCA
GCTTTATCAGAAATAATGGCAAAGTTACAAGATAAGAGTCAGCTAGAGCCAATTTCTGAAGACTTA
AATTTGCAAGGCAACCTGCGAGAATATCAAAAGCGGGGTGTAGCCTGGTTACAATATTTAGAAAAA
TTGGGATTAAATGGCTGTTTAGCCGATGATATGGGACTGGGTAAAGTCCGTGCAGGTAATTGCGAGA
TTAGTACAGGAGAAAGATAGCCAAAGTTCCCCATTACCGACATTATTAATTGCGCCGACTTCGGTT
GTTGGTAACTGGCAAAGAGAAATTGCTAAGTTTGCACCCCATTTAAAAACTATGGTGCATCATGGT
AGCGATCGCCTGCAAGATGCTGCGGAGTTTAAGTCCGCCTGTCAACAGCATGATGTGGTGATAAGT
TCCTTTACTTTGGCTCGCTTAGATGAAAACTCCTAAATAGTGTGACATGGCAACGGTTAGTTTTA
GATGAAGCACAAAACATTAAAAATCCCAAAGCAGCGCAGACTAAAGCTATACTCAAACCTCAGTGCT
AAACACCGTCTAGCTTTAACTGGTACACCAGTTGAGAACCGCTTACTTGATTTGTGGTCAATTTTT
AATTTTCTCAATCCCGGTTATTTAGGGAAAGAAGCACAGTTTCGCAAATCCTTTGAAATTCCCATC
CAGAAGGACAACGATAAAGTAAAATCGACTACCTTAAAGAACTGGTTGAACCGTTAATTTTACGA
CGGGTCAAAACAGACCAATCAATTATTAAAGACTTACCAGATAAAGTTGAACAAAAACTCTATACC
AACCTCACCAAAGAACAGGCTTCGCTATATGAAGTGGTAGTCAGAGATGTGGAAGAAAAATTGCAA
GAAGCTGAGGGAATACAACGCAAAGGTTTAATTCTCTCAACGCTGATGAAATTAAAACAGATTTGC
AATCATCCCAGACAGTTCTCTCAAGATAATAGCGAATTTTTTACCGGAGCGCTCGCACAACTTTCC
CGCTTAGTCGAAATGGTAGATGAAGCCATTTCTGAAGGAGAAAGTCTTTTAATATTTAGTCAATTT
ACAGAAGTCTGCGAACAAATAGAAAAATATCTCAAACACAACCTTACATTGCAATACCTACTACCTA
CATGGGGGTACAAGTCGCCAACGTCGGGAACAAATGATTAGTGACTTTCAAATCCTGATACGGAA
GCATCTGTATTTGTCCTTTCCCTAAAAGCTGGCGGCGTGGGGATTACTTTAACTAAAGCCAACCAC
GTCTTTCATTTTGACCGTTGGTGAATCCAGCCGTTGAAGACCAAGCCACAGACCGCGCTTTTCGC
ATAGGTCAGAAAAAAAATGTGTTTGTACATAAATTTGTGCGCCCTTGGGACTTTAGAAGAAAGAATC
GACCAAATGATTGAAGATAAGAAAAAACTTTCTTCCGCCGTAGTTGGTAGTGATGAATCGTGGCTA
ACCGAATTAGATAACGAAGCCTTTAAGAACTAATTGCCTTGAATAAAAGCACAAATTATGGAGTAG

SEQ ID NO: 62, Nostoc sp. PCC7120 Nos_sp_PCC7120_SNF2 translated polypeptide\II

MKVLHGSWIPNQYSDFVQSGAFYLVVETPINNKKRTHQTQVHPGHLSSLELLNFLTQTLGIKETEAO
LKQRICSKYFALPTANNEPLPSPELVKYLEVEVP EEYENFQYWQVTCYETVTSVKAVIAINI IKLL
KDIHFLALYNASEFQLGSDLLFWYHYTQSFRQIITKDQYIPSLKYRANAATTKKKPKQPPPGFEIY
AGWEI ISEQYEANIQKYIEYMP LICVAGNSTQTDKLEFFAPETLLRHFSEYLLNNLVSKTPLTAAF
EKQIDDSLIHYCLYPQKHNP LKTH TALQEYQQWLGWKNRIIRTQAESP FHLCFQLHSPDAEQIDNW
QMQLVSSKKDPSLKLALADYWIMNSKTKAGVHKEFGKDFDTNLLNLGYAARMYPKLWQGLETDS
PTGMQLSLDEAFDFLKDSAWVLED SGFKVIVPAWYTPAGR RRAKIRLKASSGRKVAATVGESKSYF
GLDSL VQYQYELAIGEQTLPQEWELINTKAPLVHFRGQWMELDRDKMQQLLEFWQSHGDEQPQM
SLLEFMQRSAQGEDDWEIEYDAALSEIMAKLQDKSQLEPI SEDLNLQGNLREYQKRGVAWLQYLEK
LGLNGCLADDMGLGKSVQVIARLVQEKDSQSSPLPTLLIAPTSVVGWNQREIAKFAPHLKTMVHHG
SDRLQDAAEFKSACQQHDVVISSFTLARLDEKLLNSVTWQRLVLDEAQNIKNPKAAQT KAILKLSA
KHRLALTGTPVENRLLDLWSIFNFLNPGYLGKEAQFRKSFEIPIQKDNDKVKSTTLKKLVEPLILR
RVKTDQSI IKDLPDKVEQKLYTNLTKEQASLYEVVVRDVEEKLQEAEGIQRKGLILSTLMKLKQIC
NHPRQFLQDNSEFLPERSHKLSRLVEMVDEAISEGESLLIFSQFTEVCEQIEKYLKHNHLHCNTYYL
HGGTSRQRREQMISDFQNPDEASVFVLSLKAGGVGITLTKANHV FHFDRWWNPAVEDQATDRAFR
IGQKKNVFVHKFVALGTLEERIDQMIEDKKKLSSAVVGSDES WLTELDNEAFKKLI ALNKSTIME

FIGURE 10 (continued)

66/96

**SEQ ID NO: 63, Nostoc punctiforme PCC 73102 Nospu_PCC\73102_SNF2
nucleic acid sequence**

ATGGCGATTTTACACAGTAATTGGTTACTAAAAAGTCAAAAAGGTTGTTTATTTATTTGGGGAGAA
ACTTGGCGATCGCCACGAGTTAATTTTCGAGTCTAATGGATCTGGAGATATCCCCTAAATCCATTG
GCAATGACATCACTAGAGTTGAGCGAGTGGTTGGTTTCCCAGAAGATGGCCATTACCAACTTTATC
CAGCAACCCCAAATTGCCATCGCTACTACTGGGCGAACACGTAAAGCAGCCACTGCCACTGAGATA
AACTTACCAACGCATTCACAAATAATTGCCTTACCAACTTATATTCCCAGAAGAGAGTGCAGAAGGA
ACATCTGCAATTTTCCCTGTGCATTCTGCCAGCTTGAGACTAGAAACAGACTCTCCGCAATATTTG
CAACCGTGGCTAGTTGAGGGTTTTTGTCTTAACCCCGAGCGAAGCAGTAAAATTTCTCGCTGCTGTT
CCCCTGAATGCTGCTAAAGGGGAAGATGCTTTTTTTAGGAGGAGATTTACGTTTTTTGGTCGCAAGTT
TCCCGATGGAGTTTAGATTTAATCTCGCGGTGTAAGTTTTTACCAAGAATTGAACGGCAATCAGAC
GGTGCATTTGCTGCTAAATGGCAAGTACTTCTAGACAGTGCTGTAGATGGAACCTCGCCTAGAAAAG
TTTTCTGCGGATATGCCGTTGGTTTGCCGCACTTATCAGGAGGGAGTGGGGACTGGGGACTGGGGA
CTGAGGACTGGGGAGGAGTTTTCCCAATCCCTAATCCCTAATTCCCAATCCCTACTTTATGTAAAC
TTCCCTACTGAACCTCAAGAATTGTTGCTGGGATTTCTCAACAGTACGATAGATGCCCAAGTGCGA
GGGATGGTGGGTTCTCAGCCTCCAATGGAAGCTAAGGCAATGGCATCTTTACCATCTGGGGTGCGG
CAGTGGTTGCAAGGCTTGACTAGTACATCTGGTACAGTTAACGCAGATGCCATTGAAGTGGAACGA
CTGGAAGCGGCACTGAAGGCTTGGATGATGCCGCTACAATACCAATTAACCTCTTAAACTCTATTT
CGTACCTGTTTTCAACTGCGTTCTCCAGAAGCTGGCGAAACAGATTGGACATTGGCGTATTTTTCTG
CAAGCGGCTGACGATCCTGATTTTTTTGGTGGATGCGGCAACTATTTGGAACAATCCAGTTGAACGT
TTGGTTTTATGAAAATCGAACAATTGAGCAACCACAGGAAACATTTTTTGCGAGGTTTAGGGGTAGCT
TCCCGATTATATCCAGCGATCGCACCCAGTTTTGAAACCGAATATCCCCAATCTTCTCGGATCACA
CCCATGCAAGCTTATGAGTTTATCAAGGCTGTAGCTTGGAGGTTGGAAGACAGTGGTTTGGGGGTA
ATTTTGCCTCCTAGTTTAGCGAACCGCGAAGGATGGGCAAATCGTTTTGGGTTTGAAAATTAATGCT
GAAACCCCAAAGAAAAAGCAGGGACGTTTAGGGTTGCAAAGTCTGCTGAATTTCCAATGGCAATTG
GCAATTGGCGGACAGACTATTTCCAAAGCTGAGTTTGATAAACTTGTGGCTTTAAATAGTCCACTA
GTGGAAATTAACGGTGAGTGGGTAGAATTGCGGCCCCCAAGATATCAAGACAGCCCAAACATTTTTT
ACCACTCGCAAAGACCAAATGGCGCTTTCCTTGGAAGATGCCTTGCGTTTTCAGTACAGGAGATACC
CAGGTAATTGAAAAATTACCAGTGGTCAGCTTTGAGGCATCTGGGGCATTGCAAGAGTTGATTGGG
GCGCTAAATAATAATCAAGCGATCGCACCTTTACCGACACCAGTAGGCTTTAAAGGACAGTTGCGA
CCTTATCAAGAACGTGGTGCTGCTTGGCTGTCCTTCTTGGAACGTTGGGGCTTAGGCGCGTGTCTC
GCCGACGATATGGGACTCGGTAAACTATTCAGTTTATTGCTTTTTTTGCTACATCTTAAAGAACAG
GATGCACTAGAAAATTCAACACTGCTAGTTTGTCCAACCTTCTGTTTTTAGGCAACTGGGAAAGGGAA
GTCAATAAATTTGCACCAAGCCTGAAAATTTTGCAATATCACGGTGACAAACGTCCAAAAGGGAAA
GCGTTTTTTAGAAGCAGTGAAAAATCACGATTTAATCGTTACCAGCTACTCACTGCTTCATCGGGAT
ATCAAGTCATTGCAAAGTGTTCTTGGCAGATAATTGTTTTTAGACGAAGCCCAGAATGTGAAAAAT
CCAGAGGCGAAGCAGTCAAAAGCTGTGCGGCAATTAGAAGCTACATTTTCGCATTGCATTAACGGGG
ACACCAGTAGAAAATAGACTGCAAGAACTATGGTCTATTTTGGATTTTCTCAATCCAGGGTATTTA
GGTAATAAGCAATTTTTCCAGCGGCGGTTTGCCATGCCAATTGAAAAGTATGGTGATACGGCTTCT
TTGGGTCAATTACGTTTATTAGTTTCAAGCATTTTATACTGCGGCGATTAAAAAGCGATCGCGAAATT
ATTCAAGACTTGCCAGATAAGCAAGAGATGACCGTATTTTTGCGGTTTAACTGCCGACCAAGCTGCA
CTTTATCAACAAGTTGTAGAACAATCTTTAGTAGAGATAGAATCTGCTGAAGGATTGCAACGTCGG
GGGATGATTTTTGGCTTTGCTAATCAAACCTGAAGCAAATCTGCAATCATCCAGCCCAATATTTGAAA
CAGGCGACATTAGAGCAACATAATTCAGCCAACTTCTGCGGCTAGAAGAAATGTTAGAAGAAGTT
TTAGCAGAAAGTGACCGGGCTTTAATCTTTACACAATTTGCAGAGTGGGGTAAGTTACTTAAACCC
AAAAGTGTTGAATGTTAA

FIGURE 10 (continued)

67/96

SEQ ID NO: 64, Nostoc punctiforme PCC 73102 Nospu_PCC\73102_SNF2 translated polypeptide

MAILHSNWLLKSQKGCLFIWGETWRS PRVNFESNGSGDIPLNPLAMTSLELSEWLVSQKMAITNFI
QQPQIAIATTGRTRKAATATEINLPTHSQIIALPTYIPEESAEGTSAIFPVHSASLRLETDSPQYL
QPWLVEGFCLNPSEAVKFLAAVPLNAAKGEDAFLGGDLRFWSQVSRWSLDLISRCKFLPRIERQSD
GAFAAKWQVLLDSAVDGTRLEKFSADMPLVCRTYQEGVGTGDWGLRTGEEFSQSLIPNSQSLLYVN
FPTEPQELLLGFLNSTIDAQVRGMVGSQPPMEAKAMASLPSGVRQWLQGLTSTSGTVNADAIEVER
LEAALKAWMMPLQYQLTLKTLFRTCFQLRSPEAGETDWTLAYFLQAADDPDFLVDAATIWNPNVER
LVYENRTIEQPQETFLRGLGVASRLYP AIAPSFET EYPQSSRITPMQAYEFIKAVAWRLED SGLGV
ILPPSLANREGWANRLGLKITAETPKKKQGRLGLQSLNLFQWQLAIGGQTISKAEFDKLVALNSPL
VEINGEWVELRPQDIKTAQTFFTRKDMALSLEDALRFSTGDTQVIEKLPVVSFEASGALQELIG
ALNNNQAIAPLPTPVGFKGQLRPYQERGAAWLSFLERWGLGACLADDMGLGKTIQFIAFLHLKEQ
DALENSTLLVCPTSVLGNWEREVNKFAPSLKILQYHGDKRPKGKAFLEAVKNHDLIVTSYSLHRD
IKSLQSVPWQIIVLDEAQNVKNPEAKQSKAVRQLEATFRIALTGTPVENRLQELWSILDFLNP GYL
GNKQFFQRRFAMPIEKYGD TASLGQLRSLVQPFILRRLKSDREIIQDLDPKQEMTVFCGLTADQAA
LYQQVVEQSLVEIESAEG LQRRGMILALLIKLKQICNH PAQYLKQATLEQHNSAKLLRLEEMLEEV
LAESDRALIFTQFAEWGKLLKPKSVEC

SEQ ID NO: 65, Pelodictyon phaeoclathratiforme BU-1 Pelph_BU-1_SNF2 nucleic acid sequence

ATGATTGCGCTGCACATCTCCATCATTGACGGAGTCCCGCTACTCTGGAGTGAGGGAAAAAAGATC
GGGATGCTGAAGGAGTTACGCCTCGCAACGGCTGGAATCGGCATGTTTTCCCTGCTCGACAACACC
ACAAAAGAGTTTTGTGTCTGGCTGCCCTGCCGCGAGAAAAAAGCTGTCCCATCATCTCCGCTTGTC
GGCGCCATGCCCGACCTGAGTGATGAAGAGCAACTCCATGCCTTTCCGATTACCGCGCTTCGGCTG
AATTTCAACGCTCTGTTTCGAGCTTTCCCTGCTTACGGAAAAGGGCAACATCCCCGGCAGTGGCATC
ATCTTCGGAAGCTCTCTCCACTGGGCACGGCAGGTAGTAAAAAATTGCACTGAACATTGTCAGAACC
CAGTCGCTGCTCCCTTCGATCATCAAAAACGATACATTCTGGGAGGCCTTGTGGTTGCCCTCCCC
GACAGTGCCACATCCCTCGCAGTTGAACAGCTTGCCGATGCCATGCCTGCGGTCTGTGCTCTCTC
GGCCGCACCGACACGCAACCGCCGGAAACACCAAAAAAGTTACTGCTCAAAGGACTTCTCTCTTTC
CTTGTCAATACACTGTCACGTACTTTTGAAAGAGCAGGGGTGCCAAAAATCAGTGACTTCGAGAGT
ATCCATGACGCGTGGCTTCATGCATTATCAAACAGTGATCCCCGGCTGAAATGGAAAAATGAGCAG
GAGATTGAGCAGTTTGCCTGTCAGCTCAACGCATGGCGGCGTCCCATTGACCTGCATGAGCGATCA
CCCTTCAGGTTTTTGCTGCAACTGACAGAGCCACCACTGAAAGGGCGGAAAAAGGAGCGCTGGCAT
GTTGCCTATCAACTGCAGTTGAAAGCGGATCCAAGCCTGATTCTTGACGCCGGGGATCTCTGGAAC
CCCGAAAGCGAGGCATCACAGCACGCTTTAACGTATACCTCCGATTGTACCGAATTCCTGCTTACT
TCCCTGGGACAAGCCTCCGGCCTCTGCCCCGAGTCACCCAAAGCCTGAAAAAGAAGCAGCCGGGT
GGCTTTGATCTTGATACCGAAGGGGCTTACAGATTTTTTGCTGGAGTATGCGGAACTGTTGCGAAGC
GCAGGATTTGTGGTCAAGCTTCCCTCGTGGTGGATCGGTCGCAGAGGAGTCAACCGTATCGGGATC
AAGACAAAAGTGAAGCTTCCCTCTATGAAAGGAAGCGGGTCTCACGCTGGATCGCATGGTT
GCCTGCGATTATGCTGCTGCACTTGGCAATGAGGAGCTTGACCTGCAGGAGCTGAAAACACTGGCA
AACCTGAAAGTTCCGCTGGTACGGGTGCGCGGACAGTGGACACAGATTGACCATAAGGAGCTTGCC
AATGCTCTCCATTTTCTTGAAAAACATCCAACCTGGTGAACCTTCTGCCAGAGAACTCCTCTCAACA
GCTCTCGGAGCACAAAAAAGGAGGATGCTCTCTTTCTTCGATCGGTTGAAATCGAGGGGTGGCTT
CAGGAACTGCTTGAAAAACTTTCTCTCAGGGACAATTTGAACTGCTTCCACCACCTGAGCATTTT
GAGGGAACGCTTCGCCTCTATCAGGAGCGAGGCTTTTTCATGGCTCTCATTTCTCCGCAAGTGGGGA
CTGGGCGCCTGTCTTGCCGACGACATGGGCCTTGGCAAAACCATTCAGACGCTTGCACTGCTGCAG
CGGGAGCGTGAACCTTGAGAAAAAAGGGCGGTGCTCCTGATCTGCCCCACCTCTGTAGTCAACAAC
TGGCGAAAGGAGGCGGAGCGGTTCCTCCGATTTAGCGGTGCTGGTGCATCATGGTATCGACCGG

FIGURE 10 (continued)

68/96

ATGAAAACAGCAGATTTTCGCAAAGCTGCAAGCGCTTCAGCCCTTGTCATTTCAAGCTATGGATTG
TTACAGCGCGACCTTGAATTTCTGTCGAAGGTTCCCTGGGCAGGCATTATTCTCGATGAAGCGCAG
AACATCAAAAACCCCTGAGACAAAACAGTCAAAAAGCTGCCCCGAACAATCCGGGCTGATTACCGTATT
GCCCTGACCGGCACTCCCGTTGAAAATCATGTGCGGCGACCTTTGGGCACTCATGGATTTTCTCAAT
CCCGGTTTTTCTTGGAACCCAGCACTTTTTTCAAACAGAACTTCTACACGCCGATTCAAGTGGTATGGC
GACCTTGAGGCTTCAGCACGACTGAAGTCGCTGACCGGGCCCGTTTATTCTGCGCCGCATGAAAAGC
GACAAGTCGATTATTTCCGATCTGCCCCGACAAGATCGAAATGAAAGAGTATTGCTCGCTGACCAAA
GAGCAGGCATCGCTCTACAAGGCTGTTGTGCGATGAACTGCAGGAGAAAATTGAAAGCGCCGAAGGG
ATTGACCGGCGGGGCTTGTACTTGCGCTGCTGGTCAAGCTCAAGCAGGTCTGCAACCATCCGGCA
CATTTGCTTGGCGACAACCTCTGCCATTGCACATCGTTCAGGAAAAATAAAACGCCTGACCGAACTG
CTTGGCGACATCCGCGAAGCTGGCGAAAAAACGCTGCTCTTTACACAGTTTACCATGATGGGAACG
ATGCTCCAGCACTATCTTCAGGAGTTGTACGGTGAAGAGGTACTGTTTCTGCACGGTGGCGTAACC
AAAAAAGGCGGGATGAGATGGTAGAGAGCTTCCAGAAGGAAGAGGGCAGTTCACCCTCCATCTTT
ATTCTCTCACTGAAAGCCGGAGGAACGGGTCTTAACCTGACAACAGCGAACCACGTTGTTCACTTT
GACCGATGGTGGAAACCCGGCAGTAGAGAATCAGGCAACTGACCGGGCTTTCCGTATCGGGCAGCAC
AAAAACGTTGAAGTTCATAAATTTATTACGACGGGCACGCTCGAAGAGCGCATTGATGAGATGATT
GAGAAAAAACAACGGTCGCCGGCCAGGTTCTCGGAACGGGTGAGCAGTGGCTGACCGAACTGTCTG
ACAATGATCTGCGCAAGCTCATTATGCTCGGACAGGAAGCAATGGGAGAATAA

**SEQ ID NO: 66, *Pelodictyon phaeoclathratiforme* BU-1 Pelph_BU-1
SNF2 translated polypeptide**

MIALHISIIDGVPLLWSEGKKIGMLKELRLATAGIGMFSLLDNTTKEFCVWLPCREKKAVPSSPLV
GAMPDLSDEEQLHAFPITALRLNFNALFELSLLTEKGNIPGSGIIFGSSLHWARQVVKIALNIVRT
QSLGPSIIKNDTFWEALWLPLPDSATSLAVEQLADAMPAVCRSLGRTDTQPPETPKKLLLKGLLSF
LVNTLSRTFERAGVPKISDFESIHDWLHALSNSDPRLKWKNEQEIEQFACQLNAWRRPIDLHERS
PFRFCLQLTEPPLKGRKKERWHVAYQLQLKADPSLILDAGDLWNPESSEASQHALTYTSDCTEFLLT
SLGQASGLCPAVTQSLKKKQPGGFDLDTEGAYRFLLEYAELLRSAGFVVKLPSWWIGRRGVNRIGI
KTKVKLPSMKGSGSGLTLDRMVACDYAAALGNEELDLQELKTLANLKVPLVRVRGQWTQIDHKELA
NALHFLEKHPTGELSARELLSTALGAQKKEDALFLRSVEIEGWLQELLEKLSSQGQFELLPPPEHF
EGTLRLYQERGFWSLSFLRKWGLGACLADDMGLGKTIQTLALLQRERELGEKRAVLLICPTSVVNN
WRKEAERFTPDLAVLVHHGIDRMKTADFRKAASASALVISSYGLLQRDLEFLSKVPWAGIILDEAQ
NIKNPETKQSKAARTIRADYRIALTGTPVENHVGDLWALMDFLNPGLGTQHFFKQNFYTPIQWYG
DPEASARLKSALTGPFILRRMKSDKSIISDLPDKIEMKEYCSLTKEQASLYKAVVDELQEKIESAEG
IDRRGLVLALLVKLKQVCNHPAHLGDNLSAIAHRSGKIKRLTELLGDIREAGEKTLTFTQFTMMGT
MLQHLYLQELYGEEVLFLHGGVTKKRRDEMVESFQKEEGSSPSIFILSLKAGGTGLNLTTANHVHF
DRWWNPAVENQATDRAFRIGQHKNVVHKFITTTGTLEERIDEMIEKKTTVAGQVLGTGEQWLTELS
NNDLRKLIMLGQEAMGE

**SEQ ID NO: 67, *Prochlorococcus marinus* str. CCMP1375 Proma
CCMP1375 SNF2 nucleic acid sequence**

ATGACTCTGCTGCACGCCACTTGGATTTCAACTAATTGGCATCCATCTAATTTAGGTCAATCAGAA
TTGTTCCCTTTGGGCAGACCAATGGCGCGTAGTAACCTCCAAAACAAATAATACAAACACCTTCACCT
CACCCGTTTATAGCCTATCTTCAGATGAATTAAAAGAATGGCTCAATAGCAAAAAAATTATTGCCTAAT
GAGAGTATTAATACATCTGCATGTCTCACTCTTCCTAGTAAACCCATTCACAAAAAAAATAACCAA
AAATCTAAGAATCAAAAAACTGGTATTGAATCTGAATGGAAGGGACTCCCTTTACAAGCTCATGAA
GAAATAGCAACACAATATGAATGTTGGCCATGGAAAGTAGATGGAATTTCACTCACTACTGTGCGAA
GCAACAGAATGGCTTACAAAATTACCTTTATCAAAAAAAGATTCTGATCTTAGTGAAGAATTACTT
TGGTGGGCTCATTTAGAGCGTTGGTCTCTTAATCTAATTGCGAGTGGACTATGGCTACCTCAAGTT

FIGURE 10 (continued)

69/96

AAATTACACAAGAAAGAAGGAAATGAATATCGTGTCATCATGGATACCTCTGCTGAATCAAGAAAAT
GAAAGAAATCGCTTAGAAGAGTTTGCAAAAAATATTCCCTTGGTCGCTATTTGTGCAGTCCCATGG
ATAGAAGCTAAAGGACAAATAGTCAATACTGAGCAAGTCTCAAATTCAAACAATAATACACTCTCT
TTATATAGGCCAAGACACAATCGCGTAGAAGTGATGGATCTTCTCGAAGAACTTATTGATGCACAA
CTTCGAAAAGATTTTCAACCAAGAACTAAAACTTGGATCCATTGTTAAAAGCGTGGCAAGAAGCA
CTTGGCACGAAAGATGGAATAATTAACCTATCGAATGAAAACGCTAAAAGATTAGAAAAAGCAAGT
AAGAATTGGAAAAGAGGGTTGTCTAGTAATGTTCAACCTGCGAAAACATGTCTAGAGCTAATTGCA
CCGATTGATGATCTAGATTTTATGGGACTTAACTTTTCATTGCAATCAGAATCAGATCCGAGTATC
AGACTAGCTGCAGATCAAATTTGGGAAGCAGGCGTAGAAGTAACCAAAGTTGGCGGAATAACAATT
GACAACCCAAGTGAAATTCTTTTAGAAGGCCTAGGAAGAAGTCTTGAAATTTTCCCTCCAATTGAA
AAAGGACTAGAAAGCCCAACTCCTCACACAATGAAACTGTCTGCATCAGAAGCATTTGTACTTATT
AGAACAGCAGCAGCAAAACTTCGTGACATGGGTATTGGTGTAATACTGCCTAATAGTTTGTCCAAA
GGATTTGCAAGTCGACTTGGTCTTGCTATTCAAGCCGAATTACCAGAGTCTTCACTAGGCGTAATG
CTAGGAGAAAGTTTGAAGTGGGATTGGGAGTTAATGATCGGAGGTATAAATTTAAGCATGAAAGAA
CTAGAAATGCTTGCAAAAAAAATAGTCCTCTACTCAATCACAAAGGGGACATGGATCGAATTACGT
CCTAATGATCTGAAAAATGCTTCAAATTTTTTTGCTAATACTCCAGAATTAAACCTCGATAAAGCA
TTAAGGCTTAGTGCTAATAAAGGCAACACTTTTATGAACTTCCAGTACATCATTTTGAATCTGGA
CCAAGATTACAAAGTGTCTTAGAGCAATATCACCATCAGAAAGCGCCTGAACCTTTACCAGCACCT
AATGGATTCCATGGGCAATTAAGGCCTTACCAAGAAAGAGGTCTTGGGTGGCTTGCATTTCTTTAT
CGTTTTTAAGCAAGGAGCATGCTTAGCAGATGACATGGGGCTTGGTAAAACCTATTCAATTATTATGT
TTTATTCAGCACCTAAAAGTTCAAACGAGCTTACTAAGCCTGTACTCCTAATTGCGCCTACATCT
GTGCTGACAAATTGGAAAAGAGAGGCTGCCACTTTTACTCCAGAACTATGTATACATGAACACTAT
GGTAGTAAGAGACATTCTTCAATACCAAAATTACAAAATTATCTAAAAAAAGTTGACATTATGATC
ACAAGTTATGGGTTACTTTATCGAGATGGCGAGCTGCTACAAGAAATCGACTGGCAAGGAATAGTT
ATTGATGAAGCTCAAGCTATTAAAAATTCCAAATCAAAGCAAAGTATTATAACTAGAGCAATAAGC
AAAAATCTCATAAGTAATCCCTTTAGAAATTGCTTTAACAGGAACGCCAGTAGAAAATCGTATTAGT
GAACTATGGGCACTAATGGATTTCTTAAATCCAAAAGTATTAGGTGAAGAAGATTTTTTTTAATCAG
CGATACAAGTTACCGATTGAGCATTATGGCGACATCTCTTCATTAAAAGATCTCAAACACAGGTC
AGTCCTTTTATTTTAAGAAGATTGAAAACCGATCAATCTATTATTTCTGATTTGCCTCAAAGATT
GAATTAAATGAGTGGGTGGACTAAGCCAAGAGCAAGAGCTTCTATATAAACAAACGGTAGAGAAA
AGCTTAGATGAACTCGCCTCATTACCCATTGGTCAACGCCAGGGTAAAACATTGGGTCTACTTACT
CGTCTTAAACAAATTTGTAATCATCCAGCAATTGCTTTAAAAGAACTCAAGTCGAGAAGAATTTTC
TTATTAAGATCTTCAAATTAACAAAGACTGGAAGAAATACTACAAGAAGTGAAAGAATCTCATGAT
AGAGCTCTGCTCTTTACTCAATTTGCTGAATGGGGGCATTTATTGCAAGCGTACTTACAAACAAAA
TGGGAATCAGAAGTACCTTTCCTACACGGAGGCACTCCTAAAGGGAAGCGACAAGAAATGATAGAT
CGTTTTTCAAGATGATCCTAGAGGGCCAAATATCTTTTTTACTTTTCACTAAAAGCAGGAGGAGTGGGT
CTTAATCTAACTCGTGCGAATCATGTTTTTCATATTGATCGTTGGTGGAATCCAGCAGTAGAAAAT
CAAGCAACAGATCGTGCATACCGAATTGGTCAAAAAAAAAGTGTTATCGTCCATAAGTTTATAACC
ACCGGCACAATCGAAGAAAAAATCAATCAAATGATTCTCGAAAAGACTGAACTAGCAGAAAATATT
GTCGGATCAGGAGAAAGCTGGTTAGGGCAATTAAGTCTTGAAAAATTGAGTGAATTAGTTGCTTTA
GATAGCAATCCAGAATTCTAA

**SEQ ID NO: 68, Prochlorococcus marinus str. CCMP1375 Proma
CCMP1375 SNF2 translated polypeptide**

MTLLHATWISTNWHPSNLGQSEFLWADQWRVVPKQIIQTPSPHPFSLSSDELKEWLNSKKLLPN
ESINTSACLTLPSPKPIHKKNQKSKNQKTGIESEWKGLPLQAHEEIIATQYECWPWKVDGISLTTVE
ATEWLTKLPLSKKDSLSEELLWWAHLERWSLNLIASGLWLPQVKLHKKEGNEYRASWIPLLNQEN
ERNRLEEFAKNIPLVAICAVPWIEAKGQIVNTEQVSNSNNNTLSLYRPRHNRVEVMDLLEELIDAQ

FIGURE 10 (continued)

70/96

LRKDFQPRTKNLDPLLKAWQEALGTKDGIINLSNENAKRLEKASKNWKRLSSNVQPAKTCLELIA
PIDDDLWLWDLNFSLQSESDPSIRLAADQIWEAGVEVTKVGGITIDNPSEILLEGLGRSLEIFPPIE
KGLSPTPHTMKLSASEAFVLI RTAAAKLRDMGIGVILPNSLSKGFASRLGLAIQAELPESSLGVM
LGESLNWDWELMIGGINLSMKELEMLAKKNSPLLNHKGTWIELRPNDLKNASKFFANTPELNLDKA
LRLSANKGNTFMKLPVHHFESGPRLQSVLEQYHHQKAPEPLPAPNGFHHGQLRPYQERGLGWLAFLY
RFKQGACLADDMGLGKTIQLLCFIQHLKVQNELTKPVLLIAPTSVLTNWKREAATFTPELCIHEHY
GSKRHSSIPKLQNYLKKVDIMITSYGLLYRDGELLQEIDWQGIVIDEAQAIKNSKSKQSIITRAIS
KNLISNPFRIALTGTPVENRISELWALMDFLNPKVLGEEDFFNQRYKLPIEHYGDISSLKDLKTQV
SPFILRRLKTDQSIISDLPPQKIELNEWVGLSQEQELLYKQTVESLDELASLPIGQRQKTLGLLT
RLKQICNHPAIALKETQVEKNFLLRSSKLQRLEEILQEVKESHDRALLFTQFAEWGHLLOAYLQTK
WESEVPFLHGGTPKGRQEMIDRFQDDPRGPNIFLLSLKAGGVGLNLTRANHVFHIDRWNPVEN
QATDRAYRIGQKKSIVVHKFITTTGTIEEKINQMILEKTELAENIVGSGESWLGQLSLEKLSLVAL
DSNPEF

SEQ ID NO: 69, Prochlorococcus marinus str. MIT 9211 Proma MIT 9211 SNF2 nucleic acid sequence

ATGAGTCTGCTACACGCTACTTGGCTGCCAGCAATGCCAACCAGGAAGTTCGCATAATCCAGGACTA
CTCATCTGGGCTGATTCATGGAGAGTTGCAAAACCAAGCATAGTCAGCAATCAGCCTGTAATACAT
CCATTTGCCTTATCAGCAGCAGATTTACGTATTTGGCTATTGCAAAAAAAGCTTTTACCTAAAGAA
AGTATTGAATGTACAGCCTTATTAACCTCTACCTAGTAAATCTATTAAAACTCATTAGACAAAAAA
TTAAATGGAGTAACGGACTCACAAAATACTAGCGATCAACCTCAATGGAGTGGACTACCTTTACAA
GCAGGAGAGCCAGTAACATAAACAATGTGAATGGTGGCCCTGGCAAGTTGAAGGTATAGCAATCAAA
CCCAGTGAAGCTGCATCGTGGCTTGCAAACTTACCTCTCACGAAAAAAGATCCTGAGCTTAGTGAA
GAGATCCTATGGTGGAGTCATTTAGAACGTTGGTCTCTAAGTTTAATTGCTCGTGGCCTTTGGTTG
CCACAAGTTGAATTAATAACAATTGATAATATTGGAGCTAGAGCTAGGTGGAGTCCTTTACTTAAT
AACGAAAACGAGCGCAAAAGATTAGAAGAATTCTCTATCAGGCTTCCATTAGTAGCAACATGTGCC
ATAAAAAGAGAGGAACTTCTGAAGAAAATCAAAACCATATATTAAAGACTACTCCTAGGGAAACA
CTCGATGAATACGGACTTGACAGTATGTCGACCAATCAATAGTCGACTTCAAGTGGCTTATCTCTTA
GAAGAACTCGTGGATGGACAGCTAAGAAAAGATTTTGAGGAAAGTTCTGAAGACCTTGATCCATTG
CTGAAAGCTTGGCAAGAGGCATTAGGATCACATAATGGAGTCATTCGTCTTCCGTTGGAAGATTGT
GAAAGATTAGCCAAGGCAAGTAAAAATTGGAAAGAAAATTTATCAGGCAATGTTAAAGGTGCAAGA
GCATGCCTTGAGCTTTTTGCACCACTTGAAGGAGAAGATTTATGGGACTTACAATTCTCTTTACAA
GCTGAAGCAGATCCATCACTAAAGGTAGCAGCAGAAGCAGTATGGAATGCAGACTCAGCAGTTCTA
CAGATTGGTGATATTCAAATAGCGCAGCCTGGAGAAATTCTACTAGAAGGTCTTGGCAGAGCACTC
AATATCTTTCAACCAATAGAAAGGGGTCTGGAAAATGCTACTCCAAATAATATGCAACTCACACCT
GCAGAAGCTTTTGTCTAGTACGTACAGCCTCAAAGCAATTACGTGATATTGGTATTGGTGTAATA
CTACCTAGAAGTTTATCAGGAGGATTAGCAAGTCGACTAGGTATAGCTATTAAAGCAGAGTTAGCG
ACTAGTGCCAGAGGATTAACACTTCGAGAGAACTTAGAATGGAGTTGGGAGCTAATGATAGGGGGA
AGCATATTAAGCCTTAAAGATCTAGAACAACCTGGCAAGTAAACGCAGCCCTCTAGTTTCGCTATAAG
GATTCATGGCTTGAATTACGTCCAAATGATCTTAAAATCGCCGAAAAAATTCTGTAGCAATAATCCT
GAATTAAGCCTAGATGACGCATTAAGACTTACCGCAACTAAAGGGGAGACTCTAATGAAGCTTCCA
GTACATCAATTTAATGCTGGGCCAAAGCTCCAAGGCGTTTTAGAGCAATACCACCAACATACAAGT
CCTGAGCCTCTAGCTGCACCAGATGGCTTCTATGGACAACCTGAGGCCTTATCAAGAACGTGGCATA
GGATGGTTGGCTTTCTTGCATCGTTTTTAATCAAGGTGCATGTTTAGCAGATGACATGGGCCTGGGC
AAAACAATTCAAGTGCTTGCTTTTATTCAGCACTTAAAAAGTAACAAGGACCTCAAGAAACCTGTT
TTGCTAATTGCACCTACGTCAGTATTAACAACTGGAAACGAGAAGCTTATTCATTTACACCAGAG
TTATCTGTATTAGAGCATTACGGTCCTAATCGTTCATCTACATCAACACTCTTGAAAAAGATTCTC
AAAAAAGTAGACATTCTTATTACTAGCTATGGCCTACTACATAGAGATAAACAGCTTCTGAAAACA

FIGURE 10 (continued)

71/96

ATTGATTGGCAAGGTGTAATTATTGATGAAGCACAAGCTATAAAAAATCCAAATTCAAAACAAAGT
CAAACAACCTCGTGAAATTGTTAAAGGCGGAAAAATAATCCCTTTTCGTATTGCATTAACCTGGTACC
CCTATAGAAAATCGTGTAAGTGAGCTTTGGTCATTAATGGATTTTTTAAATCCATCAGTACTTGGA
GAAAAAGAATTTTTTGATCAACGCTACAAATTACCGATTGAACGTTATGGTGATATTTCTTCGTTA
ACCGATCTCAAAGCTCGTGTCAGTCCCTTTATTCTTAGAAGGTTAAAAAGTGATAAATCAATTATC
TCGGATCTACCAAGCAAAGTCGAACTAAAAGAATGGATTACTCTTAGTCAAGAGCAAAGAGCTCTT
TATAACAAAACCTGTAGACAATAACCTTACAGGAAATCGCAAGAAGTCCTATTGGTCAGCGTCATGCG
AAAACCTTAGGTCTATTAACACGTCTCAAACAAATATGTAATCATCCTGCTCTTGCCCTCAAAGAA
AAAAACATTAGCGATGATTTTTGGAATACGATCAACCAAACCTTCAAAGGCTGGAAGAACTTCTTGAT
GTGATATTCGCAACAGAGGACAGAGCTCTTCTTTTTACCCAATTCGCTGAATGGGGTCACTTACTA
CAAGCTTATCTAGAAAAAAAGTGGGGACATAGCATACTTTTTCTACATGGAGGAACTCGCAAAATA
GATAGACAATCAATGGTTGATCAATTTCAAGAAGATCCCAGAGGCCCAAAATTATTTTTTACTTTCT
CTCAAAGCAGGTGGTATTGGTCTGAACCTGACTCGAGCTAACCACGTGTTGCATATTGATCGATGG
TGGAACCCTGCCGTAGAAAATCAGGCAACAGATCGTGCTTATAGAATTGGTCAAAAAAATAGCGTA
ATGGTTCACAAATTTATTGCTACAGGGTCAGTAGAAGAAAAAATTGATCAAATGATTACTGAAAAG
TCTAAGCTCGCAGAAAATATAATTGGTGCAGGTGAAGATTGGCTTGGCAAACCTTGGCATCAATGAA
TTACGTGAATTAGTTTCCTTAGAAAAAGAGAGTTAA

SEQ ID NO: 70, Prochlorococcus marinus str. MIT 9211 Proma MIT 9211 SNF2 translated polypeptide

MSLLHATWLPAMRTGSSHNPGLLIWADSWRVAKPSIVSNQPVIHPFALSAADLRIWLLQKKLLPKE
SIECTALLTLPSKSIKNSLDKKLNGVTDSDQNTSDQPQWSGLPLQAGEPVTQKQCEWWPWQVEGIAIK
PSEAASWLANLPLTKKDPELSEIILWWSHLERWSLSLIARGLWLPQVELNTIDNIGARARWSPLLN
NENERKRLEEFsirLPLVATCAIKREETSEENQNHILKTTPRETLDEYGLAVCRPINSRLQVAYLL
EELVDGQLRKDFEESSEDLDPLLKAWQEALGSHNGVIRLPLEDCERLAKASKNWKENLSGNVKGAR
ACLELFAPLEGEDLWDLQFSLQAEADPSLKVAEAVWNADSAVLQIGDIQIAQPGEILLEGLGRAL
NIFQPIERGLENATPNNMQLTPAEAFVLVRTASKQLRDIGIGVILPRSLSGGLASRLGIAIKAELA
TSARGLTLENLEWSWELMIGGSILSLKDLEQLASKRSPLVRYKDSWLELRPNDLKIAEKFCNNP
ELSLDDALRLTATKGETLMKLPVHQFNAGPKLQGVLEQYHQHTSPEPLAAPDGFYQQLRPYQERGI
GWLAFLHRFNQGACLADDMGLGKTIQVLAFIQHLKSNKDLKKPVLLIAPTSVLTNWKREAYSFTPE
LSVLEHYGPNRSSTSTLLKKILKKVDILITSYGLLHRDKQLLKTIDWQGVIIIDEAQAIKNPNSKQS
QTTREIVKGGKIIIPFRIALTGTPIENRVSELWSLMDFLNPSVLGEKEFFDQRYKLPIERYGDISSL
TDLKARVSPFILRLKSDKSIISDLPSKVELKEWITLSQEQRALYNKTVDNNTLQEIARSPIGQRHA
KTLGLLTRLKQICNHPALALKEKNISDDFGIRSTKLQRLEELLDVIFATEDRALLFTQFAEWGHL
QAYLEKKWGHsilFLHGGTRKIDRQSMVDQFQEDPRGPKLFLLSLKAGGIGLNLTRANHVLHIDRW
WNPAVENQATDRAYRIGQKNSVMVHKFIATGSVEEKIDQMITEKSKLAENIIGAGEDWLGLKGINE
LRELVSLEKES

SEQ ID NO: 71, Prochlorococcus marinus str. MIT 9303 Proma MIT 9303 SNF2 nucleic acid sequence

ATGATTGGTTGTGGAACCTCCTGCGTGGATGGTTGCCGTTGATCGGCAGTGCACTCCTGCTCCAAGA
AACCCAACACATACTTTTTGCGTCGCGGCCATGAGCCTGCTGCACGCCACCTGGCTTCCAGCCATC
CGTACTCCGACCAGCTCCGGTCGCCCTGCGCTCCTTGTGTGGGCAGATACCTGGCGAGTCGCTACC
CCAGCAGGACCAGCAGCAACTCCCGCACTCCACCCCTTCACACTCAACCCAGACGATCTACGTGCC
TGGCTGATTGAGCGCGATCTACTGCCCGATGAAATCATCGACGCCACAGCATGTCTGACCCTGCCT
AGCCGAACAGTCAAACCGCGCAGCAAAGCCAAGAACGTATCCACTGAATCCGACGAAGACAAAGAC
CACAAAACAAGTTGGACAGGACTGCCCTTACAAGCAGGCGAACCCATTCCCAAACAGACTGAATGG
TGGCCCTGGCAGGTGCAAGGCCTGGCAGTGGAGCCTGCTGCTGCAACGGCCTGGCTTTCGAAACTG

FIGURE 10 (continued)

72/96

CCTCTTTCAGGAGATCATCCTGATCTCGCCGATGAATTGCGCTGGTGGAGCCATCTACAGCGCTGG
GCCCTGAGCATGATTGCTCGCGGACGTTGGCTACCCCAGGTGGAACCTCAGCAAGGGAGAGGGCTAT
CCCCACCGAGCACGCTGGACACCGCTACTCAACCGTGAAGATGATCGCCGCCGCTCGAAGACCTT
GCCGCTCAGCTCCCCTTAGTGGCCACCTGCGCCCTCCCCTGGCGGGAGCCCCACCGGAAGGCGTAGC
AACCGAATGACCCGCCTAAGACCAGAGGGCGATGCGAGCCGCTAACCTGTGGCTTCATGCCGACCC
CGCAGCGGTGCGCTTCGCGTAGCCAGCCTGCTGGAAGAACTCTTGATGCCCAACTGCGCACCGGA
TTTGAAGCGAGTGAGCAAGGCCTAGACCCATTGCTCACAGCCTGGCAGGAAGCACTGGGGTTCGGAC
AGCGGCGTGATCAACCTCCCCGATGAGGAAGCCGAACGTCTAGCGACAGCAAGCAACCATTGGCGA
GAAGGCGTGGCTGGCAACGTGCGACCAAGCCAGGGCCTGCTTAGAACTCTTCACTCCCGGCGAAGGG
GAAGACCTCTGGGAGCTGCGCTTCGCCTTACAGGCTGAGGCTGATCCACGATCAAAGTACCGGCC
GCAGCAGCCTGGGCAGCGGGTCCCAAGGTCTGCAACTAGGCGAAATCCGTGTGGAACATCCAGGC
GAGGTGCTACTGGAAGGCATGGGGCGAGCCCTCACGGTGTTCGACCGATCGAACGAGGCCTCGAC
AGCGCCACACCAGAAGCAATGCAGCTCACCCCTGCTGAAGCCTTTGTATTGGTGCGCACTGCAGCG
GCCCAACTGCGTGATGTTGGCGTTGGCGTGGAATTGCCTGCCAGCCTCTCGGGAGGGGCTGGCCAGT
CGCCTAGGCCTAGCGATCAAGGCGGAGCTATCGGAGAGATCTAGAGGTTTCACTTTGGGCGAAACC
CTCGACTGGAGTTGGGAGCTCATGATCGGTGGCGTCACCCTGACGCTTCGCGAGCTGGAGCGACTA
GCAAGCAAGCGCAGCCCCGCTTGTCAACCACAAGGGCGCCTGGATCGAATTACGCCCCAACGATCTC
AAAAATGCGGAACACTTCTGCAGCGTCAATCCAGGCATCAGCCTCGACGATGCCTTGCGCCTTACC
GCAACCGATGGCGACACGCTGATGAGACTGCCCGTTACCGCTTTGAGGCCGGTCCACGACTACAG
GCGGTGTTGGAGCAGTACCACCAGCAAAAAGCTCCCGACCCCCCTACCTGCTCCCGAAGGCTTCTGC
GGTCAGCTAAGGCCTTATCAGGAAAGGGGTCTGGGTGGCTGGCCTTCCTGCATCGCTTCGATCAA
GGGGCATGCCTGGCCGACGACATGGGCCTGGGCAAAACGATCCAGCTACTGGCATTCCTGCAACAT
CTCAAGGCGGAACAGGAACCTCAAACGGCCGGTATTGCTTATCGCTCCCACATCCGTACTTACCAAC
TGGAAGAGAGAGGCATTGGCCTTCACACCAGAGTTAAACGTCCGAGAACACTATGGGCCGCGTCGG
CCCTCTACCCCCGCGCCTTAAAGAAAGCACTCAAAGGCTTAGACCTCGTTCTCACCAGTTACGGG
CTCCTGCAGCGAGATAGTGAGCTCCTGGAAACGGTCGACTGGCAAGGAGTGGTCATCGATGAAGCC
CAAGCCATTAAGAACCCCAACGCCAAACAGAGCCAAGCAGCACGCGATATGGGCCGCCAGACAAA
ACAATCGCTTCAGGATTGCTCTTACCGGCACACCCGTCGAAAACCGAGTCAGTGAACCTTTGGGCA
CTGATGGACTTCCTCAACCCAAGGGTTCTCGGTGAAGAAGACTTCTTCCGCCAGCGCTACCGGCTG
CCAATTGAACGCTATGGCGACATGTCTTCCCTGCGAGACCTCAAAGGCCGTGTTGGTCCCTTCATC
CTGAGACGACTAAAAACCGACAAGGCAATCATCTCCGACCTACCTGAAAAGGTAGAGCTGAGCGAA
TGGGTGGGTCTGAGCAAAGAACAGGCAGCCCTCTATCGCAACACAGTGGATGAAACACTGGAGGCC
ATTGCCCCGCGCACCCAGTGGTCAACGTCATGGCAAGGTGCTCGGCTTGCTTACCCGACTGAAGCAA
ATCTGCAACCATCCCGCCCTAGCCCTCAAAGAAAAAACCGTTGCAAAAGGCTTCATGGACCGCTCC
GCCAAGCTGCTGCGTTTGAAGAAATTCTCGAGGAAGTGATCGAGGCAGGAGATCGCGCTCTGTTA
TTCACCCAATTTCGAGAATGGGGTCATCTCCTTAAGGCCTACCTGCAACAACGCTGGCGCTTTGAA
GTTCCCTTCCTGCACGGCAGCACAAGCAAACTGAACGTCAGGCCATGGTTGATCGCTTCCAGGAG
GATCCACGTGGACCCCAACTGTTCCCTGCTGTCACCTCAAAGCCGGTGGCGTAGGCCTAAACCTCACG
CGGGCTAGCCATGTGTTTCATGTGATCGCTGGTGGAAATCCTGCCGTAGAAAACAGGCCACTGAT
CGCGCTTACAGGATCGGACAAACCAATCGGGTGATGGTGCACAAATTCATCACCAGCGGCTCAGTT
GAAGAGAAAATTGATCGCATGATTCGCGAAAAATCTCGACTTGCCGAAGACATCATTGGCTCTGGA
GAAGACTGGTTAGGTGGCTTAGGCGTCAGTCAATTGCGCGAACTAGTGGCCCTAGAAGACAGCTGA

**SEQ ID NO: 72 Prochlorococcus marinus str. MIT 9303 Proma MIT 9303
SNF2 translated polypeptide**

MIGCGTPAWMVAVDROCTPAPRNPTHFCVAAMSLHATWLP AIRTPTSSGRPALLVWADTW RVAT
PAGPAATPALHPFTLNPDDLRAWLIERDLLPDEIIDATACLTLPSRTVKPRSKAKNVSTESDEDKD
HKTSWTGLPLQAGEPIPKQTEWWPWQVQGLAVEPAAATAWLSKLPLSGDHPDLADELRWWSHLQRW

FIGURE 10 (continued)

73/96

ALSMIARGRWLPQVELSKGEGYPHRARWTPLLNREDDRRRLDLAAQLPLVATCALPWREPTGRRS
NRMTRLRPEAMRAANPVASCRPRSGRLRVASLLEELLDLDAQLRTGFEASEQGLDPLLTAWQEALGSD
SGVINLPDEEAERLATASNHWREGVAGNVAPARACLELFTPGEGEDLWELRFALQAEADPTIKVPA
AAAWAAGPKVLQLGEIRVEHPGEVLLEGMGRLTVFAPIERGLDSATPEAMQLTPAEAFVLVRTAA
AQLRDVGVGVELPASLSGGLASRLGLAIKAELSERSRGFTLGETLDWSWELMIGGVTLTLRELERL
ASKRSPLVNHKGAWIELRPNDLKNAEHFCSVNPGISLDDALRLTATDGDITLMRLPVHRFEAGPRLQ
AVLEQYHQKAPDPLPAPEGFCGQLRPYQERGLGWLAFLHRFDQGACLADDMGLGKTIQLLAFLQH
LKAEQELKRPVLLIAPTSVLTNWKREALAFTPELNVREHYGPRRPSTPAALKKALKGLDLVLTSYG
LLQRDSELLETVDWQGVVIDEAQAIKNPNKQSQAAARDMGRPDKNNRFRIALTGTPVENRVSELWA
LMDFLNPRVLGEEDFFRQRYRLPIERYGDMSSLRDLKGRVGPFI LRRLKTDKAIISDLPEKVELSE
WVGLSKEQAALYRNTVDETLEAIARAPSGQRHGKVLGLLTRLKQICNHPALALKEKTVAKGFMDS
AKLLRLEEILEEVIEAGDRALLFTQFAEWGHLKAYLQQRWRFEVPFLHGSTSKTERQAMVDRFQE
DPRGPQLFLLSLKAGGVGLNLTRASHVFHVDRWWNPAVENQATDRAYRIGQTNRMVHKFITSGSV
EEKIDRMIREKSRLAEDIIGSGEDWLGGGLGVSQQLRELVALEDS

SEQ ID NO: 73, *Prochlorococcus marinus* str. MIT 9313 Proma MIT 9313 SNF2 nucleic acid sequence

ATGATTGGTTGTGGAACCTCCTGCGTGGATGGTTGCCGTTGATCGGCAGTGCACCTCCTGCTCCAAGA
AACCCAACACATACTTTTTGCGTCGCGGCCATGAGCCTGCTGCACGCCACCTGGCTTCCAGCCATC
CGTACTCCGACCAGCTCCGGTCGCCCTGCGCTCCTTGTGTGGGCAGATACCTGGCGAGTCGCTACC
CCAGCAGGACCAGCAGCAACTCCCGCACTCCACCCCTTCACCCTCAGCCCAGACGATCTACGTGCC
TGGCTCATTTGAGCGCGATCTACTGCCTGATGAAATCATCGACGCCACAGCATGTCTGACCCTGCCT
AGCCGAACAGTCAAACCGCGCAACAAAACCAAGAACGTATCCACTGAATCCGACGAAGCCAAAGAC
AACAAAACAAGTTGGACAGGACTGCCCTTACAAGCAGGCGAACCCATTCCCAAACAACAGAATGG
TGGCCCTGGCAGGTGCAAGGCCTGGCAGTGGAACTGCTGCCGCAACGGCCTGGCTTTCGAAACTG
CCTCTTTCAGGAAATCATCCTGATCTGGCCGATGAATTGCGCTGGTGGAGCCATCTACAGCGCTGG
GCCCTGAGCATGATTGCTCGCGGACGTTGGCTACCCCGAGGTGGAACCTCAGCAAGGGAGAGGGCTAT
CCCCACCGAGCACGCTGGACACCGCTACTCAACCGTGAAGATGATCGCCGCCGCTCGAAGACCTT
GCCGCTCAGCTTCCCTTAGTGGCCACCTGCGCCCTCCCCTGGCGGGAGCCCACCGGAAGGCGTAGC
AACCGAATGACCCGCCTAAGACCAGAGGCGATGCGAGCCGCTAACCTGTGGCTTCATGCCGACCC
CGCAGCGGTGCGCTTCGCGTAGCCAGCTTGCTGGAAGAACTCTTGGATGCCCAACTGCGCACCGGA
TTTGAAGCGAGTGAGCAAGGCCTAGACCCATTGCTCACAGCCTGGCAGGAAGCACTGGGGTCCGAC
AGCGGCGTGATCAACCTCCCCGATGAGGAAGCCGAACGTCTAGCTACAGCAAGCAACCATTTGGCGT
GAAGGCGTGGCTGGCAACGTCGCACCAGCCAGAGCCTGCTTAGAACTCTTCACTCCCGGAGAAGGG
GAAGACCTCTGGGAGCTGCGCTTCTCCTTACAGGCTGAGGCTGATCCCACAATCAAAGTACCGGCC
GCAGCAGCCTGGGCAGCTGGTCCCAAGGTGTTGCAACTAGGCGAAATCCGTGTGGAACATCCAGGC
GAGGTGCTACTGGAAGGCATGGGGCGAGCCCTCACGGTGTTTGCACCGATCGAACGAGGCCTCGAC
AGCGCCACACCAGAAGCAATGCAGCTCACCCCTGCTGAAGCCTTTGTATTGGTGCGCACTGCAGCG
ACCCAAGTGCCTGATGTTGGCGTTGGCGTGGAATTGCCTGCCAGCCTCTCGGGAGGGGCTGGCCAGT
CGCCTAGGCCTAGCGATCAAGGCGGAGCTATCGGAGAGATCTAGAGGTTTCACTCTGGGCGAAACC
CTCGACTGGAGTTGGGAGCTCATGATCGGTGGCGTACCCTGACGCTTCGCGAACTGGAGCGACTA
GCAAGCAAGCGCAGCCCGCTTGTCACCCACAAGGGCGCCTGGATCGAATTACGCCCCAACGATCTC
AAACATGCGGAACACTTCTGCAGCGTCAATCCAGGCATCAGCCTCGACGATGCCTTGCGCCTTACC
GCAACAGATGGCGACACGCTGATGAGACTGCCCGTTACCGCTTTGAGGCCGGTCCACGACTACAG
GCGGTGTTGGAGCAGTACCACCAGCAAAAAGCACCAGACCCCTACCTGCTCCCGAAGGCTTCTGC
GGTCAGCTAAGGCCTTATCAGGAAAGGGGTCTGGGTGGCTGGCCTTCTGTCATCGCTTCGATCAA
GGGGCATGCCTGGCCGACGACATGGGCCTTGCAAAACGATCCAGCTACTGGCATTCTGCAACAT
CTCAAGGCGGAACAGGAACTCAAACGGCCGGTATTGCTTATCGCTCCACGTCCGTACTCACCAC

FIGURE 10 (continued)

74/96

TGGAAGAGAGAGGGCGTTGGCCTTCACACCAGAGTTAAACGTCCGCGAACACTATGGGCGCGTCGG
CCCTCTACCCCCGCGCCTTAAAGAAAGCACTCAAAGGCTTAGACCTCGTTCTCACCAGTTATGGG
CTCCTGCAGCGAGATAGTGAGCTCCTGGAAACGGTCGACTGGCAAGGCGTGGTCATCGATGAAGCC
CAAGCCATTAAGAACCCCAACGCCAAACAGAGCCAAGCAGCACGCGATATGGGCGCGCCAGACAAA
ACAATCGCTTCAGGATTGCTCTTACCGGCACACCCGTGCGAAAACCGAGTAAGTGAACCTTTGGGCA
CTAATGGACTTCCTTAACCCAAGGGTTCTCGGTGAAGAAGACTTCTTCCGCCAGCGCTACCGGCTG
CCGATTGAGCGCTATGGCGACATGTCTTCCCTGCGAGACCTCAAGGGCCGTGTTGGTCCCTTCATC
CTGAGACGACTCAAAACCGACAAGGCAATCATCTCCGACCTACCCGAAAAAGTAGAGCTGAGCGAA
TGGGTGGGGCTGAGCAAAGAACAGGCAGCCCTCTATCGCAACACAGTGGATGAAACACTGGAGGCC
ATTGCCCCGCGCACCCAGGGGTCAACGCCATGGCAAGGTGCTCGGATTGCTTACCAGACTGAAGCAA
ATCTGCAACCATCCCGCCCTAGCCCTCAAAGAACAAACCGTTGCAAAAGGGTTCATGGACCGCTCC
GCCAAGCTGCTGCGTTTGAAGAATAATTCTCGAAGAAGTAATCGAGGCAGGAGATCGCGCTCTGTTA
TTCACCCAATTCGCAGAATGGGGTCATCTCCTTAAGGCCTACCTGCAACAACGCTGGCGCTTTGAA
GTTCCCTTCCCTGCACGGCAGCACAAGCAAACTGAACGTCAGGCCATGGTTGATCGCTTCCAGGAG
GATCCACGTGGACCCCAACTGTTCCCTGCTGCTCACTCAAAGCCGGTGGTGTAGGCCTCAACCTGACG
CGGGCTAGCCATGTGTTTTCATGTTGATCGCTGGTGGAAATCCTGCCGTAGAAAACCAGGCCACTGAT
CGCGCTTACAGGATCGGGCAAACCAGTCGGGTGATGGTGCACAAATTCATCACCAGCGGCTCAGTT
GAAGAGAAAATTGATCGCATGATTCTGTGAAAAATCTCGACTTGCCGAAGACATCATTGGCTCTGGA
GAAGACTGGTTAGGTGGCTTAGGCGTCAGTCAATTGCGCGAACTAGTGGCCCTAGAAGACAGCTGA

SEQ ID NO: 74, *Prochlorococcus marinus* str. MIT 9313 Proma MIT 9313 SNF2 translated polypeptide

MIGCGTPAWMVAVDROCTPAPRNPTHTFCVAAMSLHATWLPPIRTPTSSGRPALLVWADTWVRVAT
PAGPAATPALHPFTLSPDDLRAWLIERDLLPDEIIDATACLTLPSRTVKPRNKTKNVSTESDEAKD
NKTSWTGLPLQAGEPIPKQTEWWPWQVQGLAVEPAAATAWLSKLPLSGNHPDLADELRWWSHLQRW
ALSMIARGRWLPQVELSKGEGYPHRARWTPLLNREDDRRRLEDLAAQLPLVATCALPWREPTGRRS
NRMTRLRPEAMRAANPVASCRPRSGRLRVASLLEELLDAQLRGTGFEASEQGLDPLLTAWQEALGSD
SGVINLPDEEAERLATASNHWREGVAGNVAPARACLELFTPGEGEDLWELRFSLQAEADPTIKVPA
AAAWAAGPKVLQLGEIRVEHPGEVLLEGMGRALTVFAPIERGLDSATPEAMQLTPAEAFVLVRTAA
TQLRDVGVGVELPASLSGGLASRLGLAIKAELSERSRGFTLGETLDWSWELMIGGVTLTLRELERL
ASKRSPLVNHKGAWIELRPNDLKHAHFCSVNPGLSLDDALRLTATDGD TLMRLPVHRFEAGPRLQ
AVLEQYHQKAPDPLPAPEGFCGQLRPYQERGLGWLAFLHRFDQGACLADDMGLGKTIQLLAFLQH
LKAQEQLKRPVLLIAPTSVLTNWKREALAFTPELNVREHYGPRRPSTPAALKKALKGLDLVLTSYG
LLQRDSELLETVDWQGVVIDEAQAIKNPNAKQSQAARDMGRPDKNRFRIALTGTPVENRVSELWA
LMDFLNPRVLGEEDFFRQRYRLPIERYGDMSSLRDLKGRVGPFILRLKTDKAIISDLPEKVELSE
WVGLSKEQAALYRNTVDETLEAIARAPRGQRHGKVLGLLTRLKQICNHPALALKEQTVAKGFMDRS
AKLLRLEEILEEVIEAGDRALLFTQFAEWGHLKAYLQQRWRFEVPFLHGSTSKTERQAMVDRFQE
DPRGPQLFLLSLKAGGVGLNLTRASHVFHVDRWWNPAVENQATDRAYRIGQTSRVMVHKFITSGSV
EEKIDRMIREKSRLAEDIIGSGEDWLGGGLGVSQ LRELVALEDS

SEQ ID NO: 75, *Rhodococcus* sp. RHA1 Rho_sp_RHA1_SNF2 nucleic acid sequence

ATGGCGCGAGCAGGGACTTCACGCGCTGTGGGTGCGACCTGCTTGGATGGGTGCATGCTGCACGGC
CTCTGGACACCGGGTTCGGGTCTCATGCTGTGGGTGGAGGATCGGAATCCGGCAGCTCCGGAGCCG
ACGGACGCGGTGCGGGCGGATGCTGGCGCGGAAGTTCCGGCATCACGTGAAGGTGCCGATGCCGACG
CCGTGCGGGGCCGGAGATGCTCGAGTGGGCCGCGGTTGCGCTCGCACCAACCGGATGCGACGGAGTTC
CTGCTGTCGGTGTCTGTCGCCGACCCCCGGATCGCCGGGGATCTGCGCTACCTCGCCCACGTGCC
CGCGGTGTGAGCGGTGGGCACGGGCCGGGCGGGTGGTGCCCGAGGTACACCGGGCGGAGGGCGGC

FIGURE 10 (continued)

75/96

TGGTGGCCGCGCTGGCGGCTGCTCGGCGGTGAACGGCAGCGTGCGTGGCTCACGGAGCTGGCCGTG
GCGATGCCGCCGGTCCAGCGTCACGGCACGACCCCCCGGGCCGTGCTCGACGACATGGTCACCGAG
CTGACCGACCCCGTCCGCCGCGGTGTCCTCGAACGACGGCACCCGGACGATTCCGGCGGGCGACGTG
GATCATCCGCTGATCGACGCGCTCGTGCGGGGTGACCAGTTCGCCGAGGGCACCGCCCAGCTGTCTG
GGATCGCTGGACGGGTGGCGCGACAGCCTCAAGGTGGACGAGCCCGAACTGGTGCTGCGGGCTCCTC
GAGCCGGAAGACGTGGACGTGGAGGGGGATTGGGACCCCGGACACGGTGCTGTGGCGACTGGAGGTC
TGCCTTCGACCGGAAGGCGAAGCCCCGGTGCCGATTCCGTTGCACCGCACGGAGGCGAGTCGTCTG
CAGATCGGGGTGCGCAAGCTGACGGAGGCCGTGGCCGCCTACCCGCGACTGCAGGACGTTCCCAGT
GACCCCGACAGCCTGGACCTGATGTTGCCACCCGCCGTGGTCATCGACCTTGTGCGGGCACGGTGCG
GTGGCGTTGAAGGAGAAGGGCATCAGCCTGCTGCTGCCGCGGGCGTGAGTGTGGCGTCGCCGTCTG
ATGCGTCTGCGGGTGAGCTCGCCGAGCACTCCGGCGAGCGCGGAGAACCAGGGCCGTGCGCAAAGAC
CAGTTGGTGCAATACTGGGAGCTGGCACTCGGCGACACGGTGCTCACCGCCGCGGAGATGAAT
CGACTGGTCAACTCCAAGAGCGATCTCGTGCGGTTGCGCGGTGAGTGGGTTCGGGCGGATCAGGAG
GTGCTCTCCCGCGCCGCGCGCTACGTGGCGGAGCGGCACGCCAGCGGCGACCCGGGCCATCGTGGAC
CTGCTGAAGGACCTGATCGCGGACGATCTGTCCGATCTTCCCGTGGAGGAGGTACGGCCACCGGC
TGGGCGGCCGCGTTGCTGGACGGCGACACGAAGCCGCAGGACGTGCCGACCCCGGACGGGTGGAC
GCCACGCTGCGCCCGTACCAGAAGCGGGGGCTCGACTGGCTGGTGTTTCATGAGCCGTCTCGGCCTC
GGGGCCGTCTCGCCGACGACATGGGACTCGGCAAGACGCTGCAGTTGCTGGCGCTGCTGGCACAC
GAGAAGGCGCCACGCCCACGCTGCTGGTGTCGCCGATGTCGGTGGTTCGGCAACTGGCAGCGCGAG
GCAGCGCGCTTCGTCCCCCTCGCTGCGGGTGCTCGTCCACCACGGTCCGCAGCGGCTGAGCGGCGCG
GAGTTCACCGCCGCGGTGACACAGAGCGATCTGGTGATCACACGTATGCGCTGCTGGCCCCGCGAC
GTCGCGCACCTGAAGGAGCAGGACTGGCGGGCGTGTCGTGCTGGACGAGGCGCAGCACATCAAGAAC
GCGAAGACGTGCGAGGCGCGGGCGGGCGCGGAGCATTCCGGCGGGCGCACCGCGTCGCGCTGACCGGC
ACTCCGGTCGAGAACCGCCTCGACGAACTGCGCTCGATCCTCGACTTCGCGAACTCGGGCATCCTG
GGCTCGGAGGTGATGTTCCGCAAGCGCTTCGTGGTGCCGATCGAGCGGGAGCAGGACGAGACAGCC
GTCGCCCCGGCTCCGCGCGGTCACGTCCCCGTTCGTGCTGCGCCGGGTCAAGACCGATCCCGCGGTC
ATCGCCGACCTCCCGACAAAGTTCGAGATGACGGTGCGCGCCAACCTCACCGCGGAGCAGGCCGCG
CTGTACCGGGCGGTGGTCGACGACATGATGGCGCAGATCAAGGACAAGAAGGGGATGAAGCGCAAG
GGCGCCGTCTCGCCGCCCTGACGAACTCAAGCAGGTGTGCAACCACCCGGCACACTTCCTGCGC
GACGGGTGCGCGGTGATGCGGCGCGGACAGCACCGCTCCGGCAAGCTGGGGCTCGTCGAGGACATC
CTGGATTCCGTGGTTCGCGGACGGCGAGAAGGCGTTGCTGTTACCCAGTTCCGGGAATTCTGGCGAC
CTCGTCACCCCGTACCTCGCGGAGCGTTTCGGTACTCCCGTGCCGTTTCTGCACGGGGGCGTGTCC
AAGCAGAAGCGCGACGACATGGTGGCCTCGTTCAGGGCGACGACGGGGCCGCCGATCATGATGCTC
TCGCTGAAGGCGGGCGGGACGGGTTTGAACCTCACCGCGGGCCAATCACGTGCTCCACCTCGACCGG
TGGTGGAATCCGGCGGTGAGAACAGGCCACGGACAGGGCGTTCCGGATCGGCCAGCGGCGGGAC
GTGCAGGTGCGCAAGCTCGTGTGCGTCGGCACCCCTGGAGGAGCGGATCGACGCGATGATCGCCACC
AAGCAGGAGCTGGCCGATCTCGCCGTGCGGACGGGCGAGAAGTGGGTGACGGAGATGAGCACCGAA
CAACTGGGCGAACTGCTCCGCCTCGGTGACGAGGCGGTGGGCGAATGA

SEQ ID NO: 76, Rhodococcus sp. RHA1 Rho_sp_RHA1_SNF2 translated polypeptide

MARAGTSRAVGRTCLDGCMLHGLWTPGSGMLMWVEDRNPAAPEPTDAVGRMLARKFRHHVKVPMPT
PSGPEMLEWAVALAPPDATEFLLSVSSRDPRIAGDLRYLAHVARGVERWARAGRVVPEVHRAEGG
WWPRWRLLGGERQRAWLTELVAMPVQRHGTTTAVLDDMVTELTDPVARRVLERRHPDDSGGDV
DHPLIDALVRGDQFAEGTAQLSGSLDGWRDSLKVDEPELVLRLLPEPDVDVEGDWDPDVTLWRLEV
CLRPEGEAPVPIPLHRTEASRLQIGVRKLTEAVAAYPRLQDVPSDPDSLMLPTAVVIDLVGHGA
VALKEKGISLLLPRASVASPSMRLRVSSPSTPASAENRAVGKDQLVQYNWELALGDTVLTAEMN
RLVNSKSDLVRLRGEWVRADQEVLSRAARYVAERHASGDRAIVDLLKDLIADDLSDLPVEEVTATG

FIGURE 10 (continued)

76/96

WAAALLDGDTPQDVPTPDGLDATALRPYQKRGLDWLVFMSRLGLGAVLADDMGLGKTLQLLALLAH
EKAPTPTLLVCPMSVVGWQREARFVPSLRVLVHHGPQRLSGAEFTA AVTQSDLVITTYALLARD
VAHLKEQDWRRVVLDEAQHIKNAKTSQARAARSI PAAHRVALTGTPVENRLDELRSILDFANSIL
GSEVMFRKRFRVVPPIEREQDETAVARLRAVTS PFVLRRVKTDPAVIADLPDKFEMTVRANLTAEQAA
LYRAVVDDMMAQIKDKKGMKRKGAVLAALTKLKQVCNHPAHFLRDGSAVMRRGQHRSGKLGLVEDI
LDSVVADGEKALLFTQFREFGDLVTPYLAERFGTPVPFLHGGVSKQKRDDMVASFQGGDDGPPIMML
SLKAGGTGLNLTAANHVVHLDRWWNPAVENQATDRAFRIGQRRDVQVRKLVCVGTLEERIDAMIAT
KQELADLAVGTGENWVTEMSTEQLGELLRLGDEAVGE

**SEQ ID NO: 77, *Salinispora tropica* CNB-440 Saltr_CNB-440_SNF2
nucleic acid sequence**

GTGCTGGTTGTCCACGGGTCGTGGCGGCTCGGCATCGGGCTCGCCATCTGGGCCGAGGACAGCGCG
TCGCCGCCTCGGGCGCCGCGCCGGGCGGGCGGGCGCCCCGCGAGCGACCCACCCGTTTCGCCGCC
GGTCACCCCGTGTCTTGCAGCAGCTCTGGCCGAGGTCGCCGAGCCGACCGAGCCCGGCACGGCACTG
CTCACCCCTGCCACCCGAGCTGGTTCGCCGCTGGACTCGCCGGAGCTGGTCCGCACCGCGTTCGGTC
GAGCCGCTCCGTGGGCCGGTCACGTTGGCCGGGTGGCGGGTGCCCGCCCTGGTTTACGCCCCGGAC
GCCGCCCTGTCTGCTCTCTCCAGATCACCGCGGCCGGCGCTCTACCTGACGCCGTACCCGGTGCC
ACTCTGCGTCACCTCGCGGAGCTGGCGGCCTTCGCCGTGGACCTCGCCGCCCGTGGTTCGGGTCCTG
CCCGGCGTCCGGCCACCGAAGGAACGTGCCAGCGCCGCCTGGGCGGTGTGGCAGCCCCCTGCTCACC
GGCGTGGACGCTGGCTGGGCCCGGGCCCTCGCCCTCGCCCTGCCGCCCGCGGTCCGTGCCGCCGTC
GAGATCGATCCGGCTCCACTCGCCGTACCCGGCGGACCGGAAACGCCCGCCAACGGTGGTGTGCCG
CCGCAGGCTCGTACGAGGCGACCGACCGCAGCCGCCGGGGAACCGAGTGAAGTGGTGGTTCGAGGCG
CTCGACGCGCTCACCGACGCGGCCGTACGGGCTGCCCTCGCGGAGACCTCCCTTACCCGGGGAGCC
CGTCCGCGGGGCGCGGTTCGCGGCCTGGCTCGCGGCGCTCACCGGCCCGCGTTCGTGACTTCACCGCC
GACTCGGCGGAGCTCGACACCCTGCGCGGTGAGTTGGACGCCTGGCAGCGCGACGCTGTGGGAGGT
TCGGTCCGGGCCAGCTTCCGGCTGGTGGAGCCGCCGACGGACGGACTCTTTGAGGCGGCGGCCGGG
GGGCTGGCCGCGGCCGAGGGGTCTGGCGGGTTCGAGTTCCGGCTACAGCCGGCCGACCGAGCCGGGT
CTGCATGTTGACGCCGTGCGGATCTGGCACGAGTCGGCGGCCCTACCGGGCCCGGCCGCTCCGCAG
GAGGCCCTGCTGACCGAGTTGGGGCGGGCCAGCCGACTCTGGCCGGAGCTGAAGTCCGGCCCTGCGC
ACCGCCACTCCAGAGGCGCTGGAGCTGGACGCCCGCGGGCGCGCATCGCTTTCTACGCGACGGCGCG
CCGGTGTCTGCACGCAGCCGGGTTCGCGGTGCTGTTGCCCTCGTGGTGGCAGCGTCCGTCTCCCGG
CTCGGCGCTCGACTACAGGCCAGAGCCGTACCGCCCCGGGCACCGTCCGCCGGGGCTGGCGACGGG
GTGGGGTTGGATGCCCTGGTTCGACTACCGCTGGGAGGTGTCCCTCGGCGACCGAGCCGCTGACCGCC
GAGGAAGTGGAGTCGCTGGCCGCGCTGAAATCTCCGTTGGTCCGCCTGCGTGGGCGCTGGGTGGAG
CTGGACCCGAAACGTCTCGCCGCCGGCCTGCGGCTGCTCCGTTCCGCCGGCGAGCTGACCGTCCGC
GACCTGCTGCGGCTCGGCCTCTCCGACCCTGCTACCGACGCGCTGCCGGTGTCTGAGGTGGCGGCC
GACGGTGCCTTGGGTGACTTGCTCGCCGGAGCTGTGGAGCGGCAACTCACCCCGGTGGACGCGGTT
CCGTCTGTTCCAGGGCGTTCCTCCGCCCTACCGAGCGGCGAGGGCTGGCCTGGCTGTCTTTCTGCAG
TCCCTCGGCCCTCGGCGGGGTGCTCGCTGACGACATGGGTCTCGGCAAGACGGTACAGCTACTCGCG
TTGCTCGCTGGTGACCCGCCGGGCGTCCGTCCGACCCTGTTGGTCTGTCCGATGTCACTGGTCCGT
AACTGGCAGCGGGAGGCGGCGACCTTCACCCCGGGCGTACGGGTCCATGTGCATCACGGCGCCGAG
CGGGCCCCGCGGGGCGGCGTTCACCGCGGCGGTGGAGGCAGCGGACCTGGTCCCTACCCACCTACACG
GTGGCTGCCCCGCGATGCGGGGGAGCTGGCCGGGGTTCGACTGGCATCGGGTGGTGGTGGACGAGGCA
CAGGCCATCAAGAACGCCTCGACGCGGCAAGCCGAGGCGGTCCGGGCGTTGCCCGCCCGGCATCGG
ATCGCGGTACCGGCACCCCGGTGGAGAATCGGCTCGCCGACCTCTGGTTCGATCATGCAGTTCGCC
AATCCCGGTCTGCTCGGCCCGGCCGCGGAGTTCAAGAAGCGGTACGCCGAACCGATCGAGCGACAC
GGCGACGCGGAGGCGGCCGAGCGGCTGCGCCGATCACCGGCCCGTTTCGTGCTGCGTTCGCTCAAG

FIGURE 10 (continued)

77/96

ACCGACTCTTCGGTTATCTCCGACCTGCCAGAGAAGCTGGAGATGGAGGTGGTGTGCAACCTGACC
GCGGAACAGGCTGCCCTCTACCGTGCGGTGGTGGACGACATGATGGCCCAGATCGAGTCCAGCGAG
GGCATCGAGCGACGTGGGCTCGTGCTGGCCGCCATGACCCGGCTCAAGCAGGTCTGCAACCACCCG
GCGCACCTGCTGCGGGACAACCTCGGCGCTGGTCGGCCGCTCCGGCAAGCTGGCCCGGCTGGAGGAG
ATCCTCGACGAGGTGCTTGTGCGGGGGGAGAAGGCCCTGCTCTTCACCCAGTACGCCGAGTTCGGC
GGCATGCTGCGCGGCCACCTGTGCGGCCCGGTTTCGGACAGGAGACGCTGTTCTTGCACGGCGGCGTC
GGTAAGGCCGACCGGGACGCGATGGTGACGCGGTTCCAGTCCCCGGACGGCCCCGCGCTCTTCGTA
CTCTCGCTCAAGGCCGGTGGTACCGGTCTCACCCTGACCGCGGCCAACCATGTCGTGCACGTTGAC
CGCTGGTGGAAATCCGGCGGTGGAGGACCAGGCCACGGACCGGGCGTTCCGCATCGGGCAGCGGCGG
CGCGTTCAGGTCCGCAAGTTTGTCTGCGCCGGCACGGTGGAGGAGAAGGTGCGCCGCGCTCATCGCC
GACAAGCGTCGGCTCGCCTCGACGGTGGTGGGTGCCGGTGAGCAGTGGGTACCAGAGCTGTCCACG
GCGCAGCTGCGGGAGCTGTTCCAGCTGGAGTCCGGGGCGGTGGCCGAATGA

SEQ ID NO: 78, *Salinispora tropica* CNB-440 Saltr_CNB-440_SNF2 translated polypeptide

VLVVHGSWRLGIGLAIWAEDSASPPRAPRRAGRAPERPHFPAAGHPVLAAALAEVAEPTTEPGTAL
LTLPTRAGSPLDSPELVRTASVEPLRGPVTLAGWRVPALVYAPDAALSLLSQITAAGALPDVPGA
TLRHLAELAAFAVDLAARGRVLPGVRPPKERASAAWAVWQPLLTGVDAGWARALALALPPAVRAAV
EIDPAPLAVPGGPETPANGGVPPQARTRRPTAAAGEPGELVVEALDALTDAAVRAALAEISLTRGA
RPRGAVAAWLAALTGPRRDFTADSAELDTLRGELDAWQORDAVGGSVRASFRLVEPPTDGLFEAAAG
GLAAAEGSWRVEFGLQPADQPGLHVDVRIWHESAALPGPAAPQEALLTELGRASRLWPELNSALR
TATPEALELDAAGHRFLRDGAPVLHAAGFAVLLPSWWQRPSSRLGARLQAQSRTAPGTVAGAGDG
VGLDALVDYRWEVSLGDQPLTAEELSLAALKSPLVRLRGRWVELDPKRLAAGLRLLSAGELTVG
DLLRLGLSDPATDALPVLEVAADGALGDLLAGAVERQLTPVDAVPSFQGVLRPYQRRGLAWLSFLQ
SLGLGGVLADDMGLGKTQVLLALLAGDPPGVGPTLLVCPMSLVGNWQREAAFTPGVRVHVHHGAE
RARGAAFTAAVEAADLVLTTYTVAARDAGELAGVDWHRVVVDEAQAIKNASTRQAEAVRALPARHR
IAVTGTPVENRLADLWSIMQFANPGLLGPAAEFFKKRYAEPIERHGDAAEAERLRRITGPFVLRRLK
TDSSVISDLPEKLEMEVVCNLTAEQAALYRAVVDDMMAQIESSEGIERRGLVLAAMTRLKQVCNHP
AHLRLDNSALVGRSGKLARLEEILDEVLVAGEKALLFTQYAEFGGMLRGHLSARFGQETLFLHGGV
GKADRDAMVTRFQSPDGPALFVLSLKAGGTGLTLTAANHVVHVDRWWNPAVEDQATDRAFRIGQRR
RVQVRKFVCAGTVEEKVAALIADKRRLASTVVGAGEQWVTELSTAQLRELFQLESGAVAE

SEQ ID NO: 79, *Symbiobacterium thermophilum* IAM 14863 Symth_IAM14863_SNF2 nucleic acid sequence

ATGATCACGGTTCACGGCAGTTTTCGTCCCCTCCGGCGCGTCCGGCTTCTTCTTCCTGTGGGGCCTG
GACGGCGTGGCCGCCCGGGATGCCGCTCCTCCCGGCCGGCGCCGCGGGGTTCGCGGCCACCCA
TGCGCAACCGAGCCGGAAGCGCTCTACCCCGCCCTGAGAGGATTGCCCTACCTGAACACCCTGTCC
CTGGTCCAGTGGCAGCCCGGACCGGACGGCGTCAGCCCGGCCCGGGTCCCGGGGATCGCCCTGTCC
GTGCCCAACGCCGTGCAGTGGCTGTTGGATCTGCCCCGACCACTTCCGCGGCACGCCCCCTCCGGCCG
GGGCACAGCCTGCAGCTCTGGTGCCTCGCATCCAAGCTGCTTCTGGAGTTCCTGGGGCGGGGCCTG
ATGCTGCCGGTGCTGCAGGCCGAGGCCGGGGTGCTGAGCGCGGGCTGGGCGCTCCACCTGACCGAC
GCCGACGACGTCCGCCGCTGACCCGGCTGGCCGCTGGATTGCCGGAGGCCTGCCGCGCCCTTGTG
CCCCCGACCGAACCCCCAACACCTACCCCTGCCGGTCGCCGACGGCCTGGTCCACCAGTTCATG
CGTACGGCGGGCCGCCGGCGTGATCCGGCTCCTCCTGGAGGAAGAGCCCCTGCCCGAGGCCAGTCG
CTACAGGATAACGCCCTGCGCCACTGGCTGGCGGCGCTGACCGGGGCGGAGGCCCGGGACCTGCCG
CCGGGCCTGCCCGGCGCGCAGGAGCTGTACGCCGCCCTGGACCGCTGGAGCGCCCCCGCCACCGGC
GTGCTGAGCCACGCCAGTCTGCGGACGGGGGTCCGCCTCCACCTGCCCGGCCCGAGACCGACGGC
GAGTGGGAGCTGGAGCTCACGCTCCATGCGCCGGACGAGGGTGCGCTGCCCGTCACCGCCGATGCC

FIGURE 10 (continued)

78/96

GTCTGGGCCAGCCTGGGCGCCGAGGTGGAGATCGGCGGGCAGCGGTACCAGGGCGCCGAGCAGCGG
CTGCTGGCCGACCTGCCGGCCATGGCCCCGCTCTTCCCGCCACTGGCGCCGCTGCTCCGGGACCCC
GCGCCCAGCCGCATGCGCATTCGCGCGGACGACGTGCTGGCCCTGATCCAGGAAGGGGCCATGCTG
CTCCAGCAGGCCGGCCACCCCGTGCTGCTGCCGGCCGCCCTTGCGAAGCCCGCCGCCCTCCGGGTC
GGAATGCGCCTCAGCCCCGCCGGGGGCAGCCCCCTCCATGTTTCGGGCTGCACCAGATCGTGAACGTG
CGCTGGGACGTGGCCCTGGGCGGCACCCCGCTCACGCTGGACGAGCTGCGCCACCTGGCGCGGCAG
AAGCGGCCCCCTGGTACAGATGCAGGGCCGGTGGGTGCGGGTGGACGAACGCACCCTGGCTGCGGTC
CTCCGCCGGATCGAGCAGCACGGCGGGCAGATGGAGCTGGGCACGGCGCTGCGCCTGGCACCCGAG
GCGGACGAGGCCACCGCGACCGGCTGGATCGCCGAGCTGCTGGAGCGGCTGCAGGAGCCAGCCCGG
ATGGAGCCGGTGCCGACCCCCGGGGGCTTCGCCGGCACCCCTGCGGGCCGTACCAGCAGCGGGGCCTC
GCCTGGCTGGCGTTCCCTGCGCCGCTGGGGCCTGGGCGCGTGCTCGCCGACGACATGGGGCTGGGC
AAGACCGTGCAGCTCATCGCCCTTCTCCTGCACGAGCGGGAGGCCGGGTGGGCCGCGGGGCCGACC
CTGCTGGTCTGCCCCGTCTCGGTCTTGGGCAACTGGTGCCGGGAGCTGGCCCGCTTCGCCCCGGGC
CTGCGGGTCCCTGGTGCACCATGGCCCCGGGAGGCTGGGCGAGCCGGACTTCGCCCCGGCAGGCCGGG
GCCCCACGACGTGGTGCTGACCACGTACTCCCTGCTGGCCCCGGGATGCCGCGCTGCTGGGCCAGGTG
ACCTGGAACGGGATCGTCGCCGACGAGGCGCAGAACCTGAAAAACCCCGACACACAGCACGCCCGG
GCGCTGCGAAGCCTTTCCGGCGGCTACCGCATCGCCCTCACCGGTACGCCCGTCGAAAACCACTG
GGCGACCTGTGGTCGCTCTTCCAGTTCCTCAACCCGGGGCTGCTGGGCAGCCGCGAGGAGTTCGAG
CGGCGCTACGCCGTGCCGATCCAGCGGTACCAGGACGAGGAGGCTGCGGGCCCGGCTCCGCCGGCAG
GTGGGTCCCTTCATCCTGCGCCGGCAGAAGAACGACCCCGCCATCGCGCCGGACCTGCCCGACAAG
CTGGAGAACACCGAGCTGGTGACCCTCTCGGTGGAACAGGCGGCGCTGTACGAGGCCATCGTGCAG
GAGACGCTGGAGCGGGCCGCGCAGGCCGACGGCATCCAGCGGCAGGCGGCGGTCCTGGCAGGCCTC
ACGCGGCTGAAGCAGGTGTGCAACCATCCCGCAGCCGCCACCGGCGACGGCCCCCTGGTGGGGCGG
AGCGGCAAGATCGACCGGCTGGTGCAACTGCTGCAGGAGGTGCTGGCGGCGGGCGAGCAGGCCCTG
CTCTTACCCAGTTCGCCCCGCTTCGGCGGGCGGCTGCAGGCCTACCTGGCGGAGACGCTGGGCTGC
GAGGTGCTCTTCTGACGGCGGCACGCCCCAGCCCGAGCGGGACCGGCTCGTCGCCCCGGTTCCAG
GCCGGCGAGGCGCCCCCTCTTCATCCTCTCGCTGAAAGCCGGCGGCCTTGGCCTCAACCTCACCGCC
GCGACCCACGTCTTTACGTGGACCGGTGGTGAATCCGGCGGTGGAGGATCAGGCCACAGACCGG
GCCTACCGCATCGGCCAGACGCGCAGGGTGCTGGTGCACCGGCTGATCACCGCCGGCACGCTGGAG
GAGCGCATCGACCGGCTGCTGGCCGAGAAGCGTGCCCTGGCGGGCCAGGTGATCATCAGCGGCGAG
TCGTGGCTCGGCCAGCTCTCCACCGAGGAGCTGCGGGCCCTGATCGCCCTGGACCGGGAGGTGTAG

**SEQ ID NO: 80, *Symbiobacterium thermophilum* IAM 14863
Symth_IAM14863_SNF2 translated polypeptide**

MITVHGSFVPSGASGFFFLWGLDGVAARDAAPPGRRRRGVPRHPCATEPEALYPALRGLPYLNTLS
LVQWQPGPDGVSPARVPGIALSVPNVQWLLDLPDHFRTPLRPGHSLQLWCVASKLLLEFLGRGL
MLPVLQAEAGVLSAGWALHLTDADDVRRLTRLAAGLPEACRALVPPDRTPNITYPLPVADGLVHQFM
RTAAAGVIRLLLEEEPLPEAQSLQDTALRHWLAALTGAEARLPPGLPGAQELYAALDRWSAPATG
VLSHASLRTGVRLHLPGETDGEWELELT LHAPDEGALPVTADAVWASLGAEVEIGGQRYQGAEQR
LLADLPAMARLFPPLAPLLRDPAPSRMRI PADVDLALIQEGAMLLQQAGHPVLLPAALAKPAALRV
GMRLSPAGGSPSMFGLHQIVNVRWDVALGGTPLTLDELRLHARQKRPLVQMQRWVRVDERTLA AV
LRRIEQHGGQMELGTALRLAPEADEATATGWIAELLERLQEPARMEPVPTPGGFAGTLRPYQQRGL
AWLAFLRRWGLGACLADDMGLGKTVQLIALLLHEREAGWAAGPTLLVCPVSVLGNWCRELARFAPG
LRVLVHHGPGR LGEPDFARQAGAHDVVLT TYSLLARDAALLGQVTWNGIVADEAQNLKNPDTQHAR
ALRSLSGGYRIALTGTPVENHLGDLWSLFQFLNPGLLSREEFERRYAVPIQRYQDEEAAARLRQ
VGPFILRRQKNDPAIAPDLPDKLENTELVTLSVEQAALYEAIVQETLERAAQADGIQRQA AVLAGL
TRLKQVCNHPAAATGDGPLVGRSGKIDRLVQLLQEVLAAGEQALLFTQFARFGGRLQAYLAETLGC
EVLFLHGGTPQPERDRLVARFQAGEAPLFILSLKAGGLGLNLTAATHVFHVDRWWNPAREDQATDR
AYRIGQTRRVLVHRLITAGTLEERIDRL LAEKRALAGQVIISGESWLGLSTEELRALIALDREV

FIGURE 10 (continued)

79/96

SEQ ID NO: 81, *Synechococcus* sp. WH 5701 Syn_sp_WH5701_SNF2
nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTGTCGGCCGACACCGCCGCGCTGCCCGCCCTGGGAGGCGGC
TACCGGCCGGGCTTGCTGCTCTGGGCCGACACCTGGCGGGTGGCGGAACCCAGACACCGGCCAGC
GAGGCGCCCCAGCACCCCTCAGCCTCGACCAGGACGACCTCGGCGCCTGGCTTGAGGAGGCCGAC
CTCTGGACGGAGGATTTCCGCCCGGCCGGAGCCACCCTCTGCCTGCCAGCCGCCGCCAGGGGGCC
AGGGGGAAAAAGAAAAGCGACACCAGCAGCTGGAGCGGCCTGCCCTGCAGGCGGGCGAGCCGATC
CCGAAATCCGTGGAGTGGTGGCCCTGGCGGGTGGAGGGCTGGTGGCTGGAGCCCGGCGCCGCCACC
CTCTGGCTTGGGCGCCTGCCCTCTCAGGCGACCATCCCGACCTGGCCGATGACCTGCGCTGGTGG
AGCCATCTGCAGCGCTGGTCGCTGAGCCTGCTGGCCCGGGGCGGCTGCTGCCCCAGGTGGAGGGG
GGCCGCGCCCGCTGGCTGCCGTTGATCAACCGCGAAGACGACCGGCGCCGCTGGAGGATCTGGCC
TCGCGTCTGCCCCAGGTGGCGGTGGCGGCCCTGGAGCCCGGCCAGGGGGAGGCCGGCGTCGCGATG
GCGTGCTGGCGGCCGGGATCCGGGCGTCGGCGGCTGGCCTCGATCCTCACGCACCTGGTGGATGCA
CGCATGCGTGCGGGCTTCACCCCCAGCGAAGAGGGGCTGGATCCGCTGCTGGCGGCCTGGCAGCGG
GCCCTCGGCCCCGGTGACGGCCGCCTCGATCTCGGGGACGACGACTGCGAACGCCTGCAGGTGGCC
ACTCACCACTGGCGCGAAGCGGTGGCTGGCCGGGTCGAGCCGGCCCGGGCCTGTCTTGAGCTCGAC
ACACCCGATGAGGGGGAAGATCTCTGGCCCCTGCGCTTCAGCCTCCAGGCCGAGGCCGATCCCAGT
CTGCTGCTGCCCGCAGCCGGGGTCTGGGGCCGCGGGGCGGCTGCCTGCAGCTGGGTGAAACCGAA
CTCCAGCAACCCGGTGAAGTCTGCTGCTGGAAGGCCTCGGGAGAGCCCTGCAGGTGTTTCGAGCCGATC
GAGAGGGGTCTCGACACCGCCACACCGGAGCGGATGGCTCTCACCCCGGCCGAAGCCTTCGTGCTG
GTGCGCACCGCCGCGCTGAAGCTGCGTGATGTGGGCGTCGGCGTGGTCCTGCCCCCAGCCTCAGC
GGTGGCCTGGCCAGCCGGCTCGGCCTCTCGATCGAGGCCGATCTGCCCGAGCGCTCCCGCGGCTTC
AGCCTCGGTGAAAGCCTGCAGTGGAGCTGGGAGCTGATGATCGGCGGCGTCACGCTCACCTGCGG
GACCTGGAGCGGCTGGCGGGCAAGCGCAGCCCGCTGGTGCAGCACAAGGGGGCCTGGATCGAGCTG
CGTCCGGGTGATCTGCGCAATGCCGAGAAGTTCTGCGCCCTCGATCCGGTCCTCAGCCTCGATGAC
GCCCTGCGCCTGACCGGCAACGAGGGGGAGACCCTGCAGCGGCTGCCGGTGCACCGCTTCACAGCC
GGCCCGAGGCTGAAGGCGGTGCTGGAGCAGTACCACCAGCAGAAGGCCCCCCGATCCCCTGCCGGCC
CCCGAGGGCTTCGCCGGCCAGCTGCGGCCCTACCAGGAGCGCGGCCTGGGCTGGCTGGCCTTCCTG
CACCGCTTCGATCAGGGGGCCTGCCTGGCCGACGACATGGGCCTGGGCAAGACAATCCAGCTGCTG
GCCTTCCTGCAGCACCTCAAGGCGGAGCAGGAAGTGAAGCGTCCCGTACTGCTGGTGGCCCCCACC
TCGGTGCTCACCAACTGGCTGCGGGAAGCGAAGGCCTTCACGCCGGAAGTGAACGTGGTGGAGCAC
TACGGCCCCCGGCGGCCCTCCACCCCCGCGGCCCTGAAGAAGAAGCTGGAGGGGATGGATCTGGTG
CTCACAGCTACGGCCTGCTGCAGCGCGACAGCGAGTTACTGAGCAGCCTCGACTGGCAGGGGGTG
GTGATTGATGAGGCCAGGCGATCAAGAATTCCTCAGCGCGCCAGTCGCAGGCAGCCCGCGATCTG
GCACGCCCCGCTCAAGCAGAGCCGCTTCCGTATCGCACTCACCGGCACCCCGGTGGAGAACCGGGTC
AGTGAGCTCTGGGCCCTGATGGACTTCCTCAATCCGAAGGTGCTTGGGGAGGAGGAGTTCTTCCGC
CAGCGCTACCGCCTGCCGATCGAGCGCTATGGCGACATGGCCTCGGTGCGCGACCTCAAGGCCCGC
GTCGGCCCCGTTTCATCCTGCGGCGCCTCAAGACTGACCGCTCGATCATCTCCGACCTGCCCGAGAAG
GTGGAAGTGAAGGAGTGGGTTGGACTCTCACCCGAGCAGGTCAAGCTCTACCGCCGCACCGTGGAG
GACACCCTCGATGCGATCGCGCGGGCACCCGTGGGCCAGAAGCACGGCCAGGTGCTGGGGCTGCTC
ACCAAGCTCAAGCAGGTCTGCAACCACCCGGCCCTGATGCTCAAGGAAGGGGAGGTGGGGGCCGGC
TTCAGCGCCCCGCTCGGCCAAGTTGCAGCGGCTCGAGGAAATCGTCGAGGAGGTGATCGCGGCCGGC
GATCGGGGCCCTCCTGTTTACCCAGTTCGCCGAATGGGGCCACCTGCTCCAGACCCACCTGCAGCAG
CGCTTCCACCAGGAGGTGCCCTTTCTCTATGGCAGTACCAGCAAGGGGGAGCGTCAGGCGATGGTG
GATCGCTTCCAGGACGACCCCCGGGGACCACAGCTGTTTCCTGCTCTCGCTCAAGGCAGGCGGCGTG
GGGCTCAACCTCACCCGGGCCAGTCATGTGTTCCACATCGACCGCTGGTGGAAATCCGGCGGTGGAG
AACCAGGCCACCGACCGGGCCTACCGCATCGGCCAGACCAACCGGGTGATGGTGCACAAGTTCATC
ACCAGCGGCTCGGTGGAGGAGAAGATCGACCGCATGATCCGCGAAAAGGCCCGCCTGGCCGAAGAC
ATCGTCGGCAGCGGTGAGGAGTGGCTCGGAGGCCTCGATCCCGGCCAGCTGCGCGACCTGGTGGCC
CTGGAGGAGTGA

FIGURE 10 (continued)

80/96

SEQ ID NO: 82, *Synechococcus* sp. WH 5701 Syn_sp_WH5701_SNF2 translated polypeptide

MSLLHATWLSADTA AVPALGGGYRPGLLLWADTWRVAEPQTPASEAPQHPLSLDQDDLGAWLEED
LWTEDFRPAGATLCLPSRRQGARGKKKSDTSSWSGLPLQAGEPIPKSVEWWPWRVEGWWELEPGAAT
LWLGRPLPSGDHPDLADDLRWWSHLQRWSLSLLARGRLLPQVEGGRARWLPLINREDDRRRLEDLA
SRLPQVAVAALEPGQGEAGVAMACWRPGSGRRRLASILTHLVDARMRAGFTPSEEGLDPLLA AWQR
ALGPGDGRDLGDDDCERLQVATHHWREAVAGRVEPARACLELDT PDEGEDLWPLRFSLQAEADPS
LLLPAAGVWAAGAGCLQLGETELQQPGELLLEGLGRALQVFEP IERGLDTATPERMALTPAEAFVL
VRTAALKLRDVG VGVVLP PSLSGGLASRLGLSIEADLPERSRGFSLGESLQWSWELMIGGVTLTLR
DLERLAGKRSPLVQHKGAWIELRPGDLRNAEKFCALDPVLSLDDALRLTGNEGETLQRLPVHRFTA
GPRLKAVLEQYHQKAPDPLPAPEGFAGQLRPYQERGLGWLAFLHRFDQGACLADDMGLGKTIQLL
AFLQHLKAEQELKRPVLLVAPTSVLTNWLREAKAFTPELNVVEHYGPRRPSTPAALKKKLEGMDLV
LTSYGLLQRDSELLSSLDWQGVVIDEAQAIKNSSARQSQAARDLARPLKQSRFRIALTGTPVENRV
SELWALMDFLNPKVLGEEFFRQRYRLPIERYGDMASVRDLKARVGPFILRRLKTDRSIISDLPEK
VELKEWVGLSPEQVKLYRRTVEDTLDAIARAPVGQKHGQVLGLLTKLKQVCNHPALMLKEGEVGAG
FSARSAKLQRLEEIVEEVIAAGDRALLFTQFAEWGHL LQTHLQQRFHQEV PFLYGSTSKGERQAMV
DRFQDDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWWNPAVENQATDRAYRIGQTNRMVMVHKFI
TSGSVEEKIDRMIREKARLAEDIVGSGEEWLGGLDPGQLRDLVALEE

SEQ ID NO: 83, *Synechococcus* sp. BL107 Syn_sp_BL107_SNF2 nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTTCCCGCCATTCGTACTTCCAGCAGTTCCGGACAACCGGCA
CTGCTCGTTTGGGCTGACACCTGGCGTGTCGCCTCACCGGAGGGACCTGGACTCACACCCGCTCTG
CATCCCTTCACCCTTGGCTCGAACGATCTCAAGGCTTGGTTGACCGAACGGGACCTGATGCCTGGG
GGCAGCATCGATGCCACCGCCTGCCTCACCTCCCAAGCCGCACCGTCAAACCCCGCAAAAGTCGA
ACCCAATCGAGCGAACCAGATCCGGAGGGGCCAGCCTGGACCGGGTTGCCAATGCAAGCGGGAGAA
CCCATTCCAAACAAATGGAATGGTGGCCATGGCAAGTGCAAGGCCTGGCGGTCGAGCCATCGGCC
GCCACGGAATGGCTGGCCCGTTTACCCCTATCGGGCCGACATCCAGACCTTGGGGATGAACTGCGC
TGGTGGAGTCACCTCCAACGTTGGTCCCTCAGCTTGGTGGCCCGTGGTCGCTGGATTCCCCAAATG
GAATTAAGCAAAGGCGAGGGGTACCCCCACCGAGCGCGCTGGGTTCCTCTGCTGAACCGTGAGGAG
GATCGACGCCGGCTCGAAGACCTCGCCGCGACGCTGCCCTCGTAGCGACCTGTGCCCTCCCTTGG
CGTGAGCCACTCGGACGCCGCGAGCAACCGCACCAAGGCTTCGACCGGAAGCGATGCGAGCCGCC
AATCCGGTCGCCTGCTGTCGCCCACGAAGCGGTGCGCTCAGGGTGGCCACCTTGCTTGAAGACTTG
GTGGATGCGGAGCTGCGCAAGGGATTTGAACCAAGCACGGAAGGCCTCGACCCCTTACTCACCTTG
TGGCAAGAGGCCCTGGCCTCAGAAACCGGTGTTGTGGAGGTGGGCAACGAAGACGCAGAACGCCTC
ACCGCGGCAAGCCTGCACTGGCGCGAGGGAATTGCCGGAGGCTTCGCGGCCGCCCCGCACCTGCCTC
GAACTCAACACCCCAAACGAAGGCGAAGAACTCTGGGACCTGAAGTTTGGATTGCAAGCGGAGGCC
GATCCCAGCCTCAAGCTGCCGGCCGCCGCGGCCTGGGCCTCAGGAGCGGAAACCTTCAACTGGGG
GAAATCCAAGTTGACCAGGCGGGGGAAGTGCTGCTGGAGGGTCTTGGCCGAGCCCTCACGGTGTTC
CCTCCGATCGAACGCGGACTGGAAAGCGCAACACCGGAAACGATGCAGCTCACTCCAGCGGAGGCA
TTTGTGTTGGTGCGAACAGCAACGCACCAAGCTCCGCAATGCCGGCATCGGCGTCGAACTGCCCCC
AGTCTTTCAGGGGGCCTCGCCAGCCGGCTTGGCTTAGCGATTAAAGCGGATCTACCGGATCGATCC
AGCGGCTTCACCCTCGGCGAATCTCTTGACTGGAGCTGGGATCTCATGATCGGCGGCGTCACACTC
ACCCTCCGAGAGCTCGAACGTCTCAGCGGTAAGCGAAGTCCGCTGGTACGCCACAAGGGCGCCTGG
ATCGAACTACGGCCCAACGATCTCCGCAACGCCGAACGCTTTTGTGGAGCCAATCCAGAACTGAGC
CTCGACGACGCACTACGGCTCACGGCCACAGAAGGGGAGCTCATGATGCGCCTGCCGGTGCATCGC
TTTGATGCAGGGCCTCGTCTTCAGGGAGTTCTCGAGCAATACCACCAGCAAAAAGCCCCCGATCCC
CTGCCAGCTCCAGAGGGATTTTCCGGACAACCTCCGTCCCTATCAAGAACGTGGCTTGGGCTGGCTG

FIGURE 10 (continued)

81/96

GCCTTCCTGCATCGCTTCGATCAGGGCGCCTGCCTGGCGGACGACATGGGCTTGGGCAAGACCATC
CAGTTATTGGCGTTTCCTGCAGCACCTCAAAGCGGAAAACGAACTCAAACGCCCCGGTGCTGTTGGTG
GCCCCAACCTCGGTGCTCACGAATTGGCGACGGGAAGCGGAAGCCTTCACCCCTGAGCTGTCCGGTG
AGAGAGCACTACGGGGCCACGCCGGCCTTCCACGCCGGCCGCCTTGAAAAAAGAGCTCAAAGGTGTG
GATCTGGTGCTCACCAGTTACGGACTGATGCAACGCGACAGTGAGCTGCTGGACAACCTCGACTGG
CAAGGGGTTGTGATCGATGAAGCTCAGGCGATCAAGAACCCTGGGGCAAAGCAAAGCCAAGCGGCC
CGAGACCTAGCGCGAGCCGGGAAGAGCAGCAGGTTCCGCATTGCACTCACGGGCACACCGGTGGAA
AACCGCGTCAGCGAGCTGTGGGCGCTGATGGATTTCTCAACCCCAAAGTGTTGGGTGAGGAAGAC
TTTTTTTCGTCAGCGCTACCGCATGCCAATTGAGCGCTACGGCGATATGTCGTCGTTACGCGATCTC
AAAGCACGGGTTGGTCCCTTCATCCTGCGCCGCCTCAAACCCGACAAGTCGATCATTTCCGACCTG
CCTGAAAAGGTGGAGCTCAGCGAATGGGTGGGGCTCAGCAAAGAACAGAAATCGCTGTACAACAAA
ACCGTTGAAGACACCCTCGATGCCATTGCCACCGCACCTCGAGGGCAACGCCATGGCCAGGTGCTG
GCGCTCTTGACCCGTTTAAACAGATTTGCAATCACCCGGCCTTAGCCCAACGCGAAGGTGCCGTT
GACGCCGAATTCCTTAGCCGGTCCGCCAAGCTCATGCGGCTGGAAGAAATCCTTGAAGAGGTGATT
GAAGCCGGCGATCGCGCTTTGCTGTTTACCCAGTTCGCCGAATGGGGACACCTCTTGCAAGCCTGG
ATGCAACAACGCTGGAAGTCTGAGGTTCCCTTTCTGCACGGCGGAACCCGCAAAAGTGATCGGCAA
GCGATGGTGGATCGATTCCAAGAGGACCCCCGGGGACCTCAACTCTTCTCTCTCCCTCAAGGCC
GGTGGTGTGGCCTAAACCTCACCCGGGGCCAGCCACGTGTTCCACGTTGGATCGCTGGTGAATCC
AGCGGTGGAAAACCAAGCCACCGACCGGGCCTATCGAATTGGTCAAACCAACCGGGTGATGGTGCA
CAAATTCGTCACCCGTGGCTCGGTGGAAGAAAAAATCGACCAAATGATTCGTGA

SEQ ID NO: 84, *Synechococcus* sp. BL107 Syn_sp_BL107_SNF2 translated polypeptide

MSLLHATWLPARTSSSSGQPALLVWADTWRVASPEGPGLTPALHPFTLGSNDLKAWLTERDLMPG
GSIDATACLTLPSRTVKPRKSRTQSSEPDPEGPAWTGLPMQAGEPIPKQMEWWPWQVQGLAVEPSA
ATEWLARLPLSGRHPDLGDELRWWSHLQRWSLSLVARGRWIPQMELSKGEGYPHRARWVPLLNREE
DRRLEDLAATLPLVATCALPWREPLGRRSNRTTLRLRPEAMRAANPVACCRPRSGRLRVATLLEDL
VDAELRKGFEPSTEGLDPLLTWQEALASETGVEVEGNEDAERLTAASLHWREGIAGGFAAARTCL
ELNTPNEGEELWDLKFGLQAEADPSLKLPAAAAWASGAETLQLGEIQVDQAGEVLLEGLGRALTVE
PPIERGLESATPETMQLTPAEAFVLVRTATHQLRNAGIGVELPPSLSGGLASRLGLAIKADLPDRS
SGFTLGESLDWSWDLMIGGVTLTLRELERLSGKRSPLVRHKGAWIELRPNDLRNAERFCGANPELS
LDDALRLTATEGELMMRLPVHRFDAGPRLQGVLEQYHQKAPDPLPAPEGFSGQLRPYQERGLGWL
AFLHRFDQGACLADDMGLGKTIQLLAFLQHLKAENELKRPVLLVAPTSVLTNWRREAEAFTELSV
REHYGPRRPSTPAALKKELKGVDLVLTSYGLMQRDSELLDNLDWQGVVIDEAQAIKNPGAKQSQA
RDLARAGKSSRFRIALTGTPVENRVSELWALMDFLNPKVLGEEDFFRQRYRMPIERYGDMSSLRDL
KARVGPFILRLKTDKSIISDLPEKVELSEWVGLSKEQKSLYNKTVEDTLDAIATAPRGQRHGQVL
ALLTRLKQICNHPALAQREGAVDAEFLSRSAKLMRLEEILEEVIEAGDRALLFTQFAEWGHLQAW
MQQRWKSEVPFLHGGTRKSDRQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHVGSLVES
SGGKPSHRPGLSNWSNQPGDGAQIRHPWLGGGRKNRPND

SEQ ID NO: 85, *Synechococcus* sp. CC9311 Syn_sp_CC9311_SNF2 nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTTCCGGCCATTCGTACTCCTACCAGCTCTGGACGAGCTGCC
CTTTTGGTGTGGGCCGACACCTGGCGCGTTGCCGAGCCTGCAGGCCCAAGTACAACCCCTGCGCTT
CACCCGTTACCCCTCAGCCCAGACGATCTCCGGGCCTTGCTCACGGAACGGGATCTTTTACCCGAC
GGCATCATTTGATGCCACGGCATGCCTCACCCCTGCCGAGCCGCAGCGTGAAGCCCCGAAAAAACGC
GAAACAGAGACCAGCAGCACTGAACAGCCCAGCTGGACAGGCCTTCCCTTACAGGCTGGAGAACCG
ATCCCCAAACAAACAGAGTGGTGGCCTTGGCAGGTTCAAGGGGCTCGCAATTGACCCCATGGCGGCC

FIGURE 10 (continued)

82/96

ACCGCCTGGCTGTCCAAACTGCCTCTGTCAGGACGACATCCTGATTTGGCTGATGAGTTGCGCTGG
TGGAGTCACATGCAGCGTTGGTCCCTCAGCCTCGTAGCCCGAAGTCGCTGGCTCCCCCAAGTGGAG
CTGAGCAAGGGCGAGGGCTATCCCCATCGCGCCCGCTGGGTACCGCTTCTGAATCGGGAAGAAGAC
AGGCGCCGTCTAGAAGACTTGGCCGCAGGGCTCCCTCTCGTTGCCACCTGTGCCCTGCCTTGGCGA
GAACCAACGGGCAAACGCAGCAACCGAATCACCAGGCTCAGACCAGAAGCCATGCGCGCCGCGAAT
CCCGTGGCTTGCTGCAGGCCTCGCAGCGGACGACTAAGGGTTGCCACGTTATTGGCCGACCTGATG
GACGCGCAGCTGCGCAAGGGCTTTACTCCTGACCCTGACGGCTTGGACCCCTGCTACGCGCCTGG
GAGGAGGCCTTGAGCTCGGATACAGGTGAAATCCAACCTCAGCGATGAAGAAACCGAACGCCTAGCC
ACCGCCAGTAATCATTGGCGTGAAGGGGTGCTGGAAATGTTGCTGCAGCCCGCGCCTGCCTGGAG
CTGGCAACACCAGCGGACGATGAGGACCTTTGGCCACTGCGCTTCTTTCTGCAGGCGGAAGCAGAT
CCAACCCTCAAGCTGCCCCGAGGAGCGGCATGGGCTGCAGGCCCCAGCGGCCTCCAACCTTGGGGAA
ATCAAGGTGGAGCACCCCGAGCGAGGTCTTGCTCGAGGGTATGGGGCGAGCCCTGACCGTGTTCCAA
CCGATCGAGCGCGGACTGGACAGTGCCACGCCAGAGAGCATGCAGCTCACACCAGCTGAAGCGTTT
GTTTTGGTGCGCACAGCAGTCCGACAACCTGCGGGATGTGGGCGTTGGCGTTGACCTGCCACCAAGC
CTGTCTGGAGGGCTGGCTAGCAGGCTTGGCCTCGCCATCAAGGCAGAACTCTCCGAGCGTTTCGCGA
GGCTTCACGCTCGGTGAAAACCTTGACTGGAGCTGGGAGCTGATGATCGGCGGGGTGACGCTGACC
TTGCGAGAGCTTGAGCGATTGGCTGGTAAGCGCAGCCCTCTGGTGCGTCACAAAGGGGGCTTGGATC
GAACTACGGCCCAATGACCTCAAAAATGCCGAGCGCTTTTGCGCCGCCAATCCAGACCTGAGCCTC
GACGACGCGCTTCGGCTCACCGCCACCGAAGGCGACACGATGATGCGCCTGCCCGTGTCATCAATTT
GATGCCGGTCCGCGGCTGCAAGCCGTGCTGGAGCAGTACCACCAGCAGAAAGCGCCAGACCCACTC
CCCGCTCCCGAGGGCTTTTCGGGTCAACTCAGGCCCTATCAAGAGAGAGGACTCGGCTGGCTTGCC
TTCCTGCATCGCTTCGACCAAGGCGCCTGCTTGGCCGATGACATGGGCCTTGGCAAACCATCCAG
CTGCTGGCTTTTCTGCAACACCTCAAGGCAGAAAACGAACTCAAGCGATCAGTGCTTTTAATTGCA
CCCACATCTGTCCTTACGAACTGGAAACGAGAGGCAACAGCGTTTACACCCGAGCTCAAGGTGCAT
GAGCACTACGGTCCAAAACGCCCGAGCACCCAGCAGCACTGAAAAAGGCGCTGAAAGACGTGGAT
CTCGTGCTCACCAGCTATGGCCTGTTACAACGCGACAGTGAGCTCCTCGAAAGTCACGATTGGCAA
GGCCTCGTGATCGATGAAGCGCAGGCGATAAAAAACCCCTCCGCGAAGCAAAGCCAAGCCGCCCGT
GATCTGGCCCGCCCGAAAAAGAACAGCCGTTTTTCGCATCGCACTCACCGGCACACCAGTTGAGAAC
CGCGTCAGCGAGCTCTGGGCCCTGATGGACTTCCTCAACCCTCGGGTACTGGGAGAGGAAGAATTT
TTCCGACATCGCTATCGCATGCCGATTGAGCGTTACGGAGACCTGTCTCTCGCTGCGCGACCTCAAA
GCCCGAGTGGGACCTTTCATCCTCAGACGACTCAAAACAGACAAAGCGATCATCTCGGATCTACCC
GAGAAGGTGGAATTGAGCGAGTGGGTGGGCTGAGCAAAGAGCAGAAGTCGCTGTATGCCAAAACC
GTTGAAGACACCTTGGATGCCATTGCCCGCGCGCCACGCGGCAAACGTCATGGTCAGGTGTTGGGT
CTGCTCACCAGCTCAAGCAGATTTGCAACCACCCTGCGCTTGCCCTCAAGGAGCAGGGCGCCAGC
GAAGATTTCTCAACGGTCCGTGAAGCTGCAACGTCTCGAAGAAATTTTGGACGAGGTTGTAGAA
GCTGGGGATCGAGCCTTGCTGTTTACCCAGTTCGCGGAATGGGGCAAGTTGCTCCAGGATTATTTG
CAACGACGCTGGCGCAGCGAAGTTCCCTTCCTCAGCGGCAGCACCAGCAAAAGTGAACGGCAAGCC
ATGGTCGATCGCTTCCAGGAGGATCCGCGCGGGCCCCAGCTTTTCCTGTTATCACTCAAAGCTGGC
GGAGTCGGCCTCAACCTCACGCGCGCCAGTCATGTCTTTCACATCGACCGTTGGTGGAACCCCGCC
GTTGAAAATCAAGCCACGGACCGTGCCTATCGCATCGGCCAAACGAACCGGGTCATGGTGATAAG
TTCATCACCAGCGGCTCCGTTGAGGAGAAAATTGACCGCATGATCCGCGAGAAGTCCAGACTGGCG
GAAGACATCATTGGCTCCGGCGAAGACTGGCTTGGAGGCCTGGAAATGGGACAACCTCAAAGAGCTA
GTGAGCCTGGAGGACAACCAAGCATGA

FIGURE 10 (continued)

83/96

SEQ ID NO: 86, *Synechococcus* sp. CC9311 Syn_sp_CC9311_SNF2 translated polypeptide

MSLLHATWLPARTPTSSGRAALLVWADTWRVAEPAGPSTTPALHPFTLSPPDDLRLALLTERDLLPD
GIIDATACLTLPSRSVKPRKKRETETSSTEQPSWTGLPLQAGEPIPKQTEWWPWQVQGLAIDPMAA
TAWLSKLPLSGRHPDLADELRWWSHMQRWSLSLVARSRWLPQVELSKGEGYPHRARWVPLLNREED
RRRLEDLAAGLPLVATCALPWREPTGKRSNRITRLRPEAMRAANPVACCRPRSGRLRVATLLADLM
DAQLRKGFTPDGDLPLLRWEEALSSDTGEIQLSDEETERLATASNHWREGVAGNVAAARACLE
LATPADDEDLWPLRFFLQAEADPTLKLPAAGAAWAGPSGLQLGEIKVEHPSEVLLEGMGRALTVFQ
PIERGLDSATPESMQLTPAEAFVLVRTAVRQLRDVGVGVLDLPPSLSGGLASRLGLAIKAELSERSR
GFTLGENLDWSWELMIGGVTLTLRELERLAGKRSPLVRHKGAWIELRPNDLKNAERFCAANPDLSL
DDALRLTATEGDTMMRLPVHQFDAGPRLQAVLEQYHQKAPDPLPAPEGFSGQLRPYQERGLGWLA
FLHRFDQGACLADDMGLGKTIQLLAFLQHLKAENELKRSVLLIAPTSVLTNWKREATAFTPELKVH
EHYGPKRSTPAALKKALKDVDLVLTSYGLLQRDSELLESHDWQGLVIDEAQAIKNPSAKQSQAAR
DLARPKKNSRFRIALTGT PVENRVSELWALMDFLNPRVLGEEFFRHRYPMPRIERYGDLSSLRDLK
ARVGPFILRRLKTDKAIISDLPEKVELSEWVGLSKEQKSLYAKTVEDTLDAIARAPRGKRHGQVLG
LLTKLKQICNHPALALKEQGASEDFLKRSVKLQRLEEILDEVVEAGDRALLFTQFAEWGKLLQDYL
QRRWRSEVPFLSGSTSKSERQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWWNPA
VENQATDRAYRIGQTNRMVHKFITSGSVEEKIDRMIREKSRLAEDIIGSGEDWLGGLEMGQLKEL
VSLEDNQA

SEQ ID NO: 87, *Synechococcus* sp. CC9605 Syn_sp_CC9605_SNF2 nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTTCCCGCCATCCGCACCTCCAGCAGTTCCGGTCAACCGGCA
CTGCTCGTTTGGGCTGACACCTGGCGGGTGGCCACACCGGAAGGCCCGGGCCTTACCCCAGCGCTG
CACCCCTTCACCCTAAGCCATGAAGACCTCAGGGCCTGGCTGAGCGAACGCGACCTCTTGCCCGGC
GGCTGCATCGATGCCACGGCGTGCCTCACCTGCCGAGCCGCACGGTGAAGCTGCGCAAAAGCCGC
AGCACAAAAGAGGAGCCAACACCGGAACCACCGGGTTGGACCGGGCTACCGATGCAGGCCGGCGAA
CCGATCCCCAAGCAAACCGAATGGTGGCCCTGGCAGGTGCAGGGGCTCGCGGTGGAACCGTCGGCA
GCCACGGAGTGGCTGTCCCGATTGCCGCTCTCCGGCACCAATCCAGACCTGGCTGATGAACTGCGC
TGGTGGAGCCATCTGCAGCGCTGGGCCTTGAGTCTGGTGGCCCCGGGGCCGCTGGATTCCCCAGATG
GAGTTCAGCAAAGGGGAGGGCTATCCCCATCGGGCCCGTTGGGTGCCGCTTCTCAACCGGGAAGAA
GACCGGCGCCGGCTGGAGGATCTGGCGGCCAGCCTGCCGCTGGTGGCCACCTGCGCCTTGCCCTGG
CGGGAACCCCTGGGGCGCCGCAGCAACCGCACCCCGGTTACGACCGGAGGCGATGCGAGCCGCC
AACCTGTGGCCAGCTGCCGGCCCCGCAGCGGACGCCTGCCGGGTGGCGACGCTGCTGGAAGATCTA
GTGGACGCGCAGCTGCGCAAGGACTTTGAACCTCCACCGATGGGCTTGATCCCCTGCTGACCCTC
TGGCAGGAGGCCCTGGGGTTCGGAGACCGGGGTGATCGAGATCGGCGATGAAGAGGCCGAACGCCTG
GCCACCGCCAGCCATCACTGGCGGGAGGGCATCGCCGGCGATTTTGCTGCGGCCCCGCACCTGCCTT
GAACTGCACACCCCAACCGGATGGGGAGGATCTCTGGGAGCTGCGCTTCGGGCTGCAGGCGGAAGCT
GACCCAGCCTGAAGCTCCCGGCCCGCGCGGCTGGGCGGCTGGTGCAGAACCGCTACAGCTTGGA
GAGATCCGGGTGGACCAACCGGGTGAAGTGCTGCTGGAAGGCATGGGCCGCGCCCTGAGCGTGTTT
CCGGCAATTGAGCGGGGTCTGGAGAGCGCCACACCTGAAACGATGCAGCTCACCCCGGCCGAGGCC
TTCGTGCTGGTGCACACGGCCCGCCCGGCAGCTGCGGGATGCCGGCGTGGGAGTGGAGCTGCCGCCC
AGCCTCTCCGGTGGCCTGGCCAGCCGACTGGGCCTGTGATCAAAGCGGAACCTGCCCGAACGCTCG
AGCGGTTTTCACGTTGGGTGAGTGTCTGGCCTGGGAGTGGGATCTGATGATCGGCGGGGTGACGCTC
ACCCTGCGGGAATTGGAGCGCCTGAGCGGCAAGCGCAGCCCCCTGGTGCGCCACAAGGGGGCCTGG
ATCGAACTGCGGCCCAACGACCTCAAAAATGCCGAACGCTTCTGTGGGGCGAAACCTGAACTGAGC
CTCGACGACGCGCTGCGGCTGACGGGGACGGAAGGGGAACTGTTGATGCGGATGCCGGTGCACCGC
TTCGACGCGCGGCCACGGCTGCAATCGGTGTTGCAGCAATACCACCAGCAGAAGGCCCCCGACCCC

FIGURE 10 (continued)

84/96

TTGCCGGCCCCGGAAGGATTCAGCGGGCAGCTGCGGCCTTATCAGGAGCGGGGCCTCGGCTGGCTC
GCCTTCCTGCACCGCTTCGATCAAGGGGCCTGTCTAGCTGACGACATGGGCTTGGGCAAACCATTT
CAGTTGCTAGCGTTCCTGCAGCACCTCAAAGCGGAGCAAGAACTGAAACGCCCCGGTGCTGCTGGTG
GCCCCCACATCGGTGCTCACCAACTGGCGACGGGAGGCGGAATCGTTCCTCCAGAGTTGAAGGTC
ACCGAGCATTACGGGCCTCGCCGGGCCCTCCACACCCGCCGAACCTCAAAAAAGCGTTGAAGGAGGTG
GATCTGGTGCTCACCAGCTACGGGCTGCTGCAGCGTGACAGCGAACTGCTGGAAACCCAGGACTGG
CAGGGGGTGTTGATTGACGAAGCCCAGGCGATCAAGAACCCTGGCGCCAAACAGAGCCAAGCCGCC
CGGGATCTGGCCCGCACCGGCCGCATCAAGAGCAACCGCTTCCGCATCGCACTCACCGGCACCCCC
GTGAAAACCGGGTGAGCGAACTGTGGGCCTTGATGGACTTCCTCAACCCAAAGGTGCTTGGGGAA
GAAGACTTCTTCCGCCAGCGCTATCGGATGCCGATTGAGCGCTACGGCGACATGTCGTCCCTGCGG
GACCTGAAAGGCCGCGTGGGTCCGTTTCATCCTGCGCCGGCTGAAAACCGACAAGACGATCATTTCC
GACCTGCCTGAAAAGGTGGAGCTGAGCGAATGGGTGGGGCTGAGCAAGGAGCAGAAATCTCTGTAC
AGCAAGACCGTGGAAGACACCCTCGATGCCATTGCCCCGGGCGCCGCGCGGGCAGCGCCACGGGCAG
GTGCTGGCCCTGCTCACCCGGCTGAAACAGATCTGCAACCATCCCGCCCTGGCCCTGAGCGAAGGG
GCCGTGGACGATGGCTTCCTGGGCCGTTTCGGCCAAGCTGCAGCGGCTGGAGGAGATCCTCGATGAG
GTGATCGAAGCGGGCGATCGGGCCCTGCTGTTACCCAGTTCGCCGAATGGGGGCATTTGCTAAGG
GCCTGGATGCAGCAGCGCTGGAAATCAGAAGTGCCCTTCCTGCACGGCGGCACCCGCAAGAACGAA
CGCCAGGCGATGGTGGATCGCTTCCAGGAGGATCCCCGCGGTCCACAGCTGTTCTCTGCTCTCGCTC
AAGGCCGGTGGTGTGGGCCTCAACCTCACGCGGGCCAGCCATGTGTTCCACATCGATCGCTGGTGG
AACCCTGCCGTGGAAAACAGGCCACCGACCGGGCCTATCGGATCGGCCAAACGAACCGAGTGATG
GTTCATAAATTCATCACCAGCGGTTCGGTGGAGGAAAAAATCGATCGCATGATCCGCGAGAAATCA
CGCCTGGCCGAAGATGTGATCGGCTCCGGCGAAGATTGGCTGGGAAGCCTCGGTGGCGATCAATTG
CGCGATCTCGTTTCTTTGGAGGACACCTGA

**SEQ ID NO: 88, *Synechococcus* sp. CC9605 Syn_sp_CC9605_SNF2
translated polypeptide**

MSLLHATWLPARTSSSSGQPALLVWADTWRVATPEGPGLTPALHPFTLSHEDLRAWLSERDLLPG
GCIDATACLTLPSRTVKLRKSRSTKEEPTPEPPGWTGLPMQAGEPIPKQTEWWPWQVQGLAVEPSA
ATEWLSRLPLSGTNPDLADELRWWSHLQRWALSIVARGRWIPQMEFSKGEYPHRARWVPLLNREE
DRRRLDLAASLPLVATCALPWREPLGRRSNRTTLRLRPEAMRAANPVASCRPRSGRLRVATLLEDL
VDAQLRKDFEPSTDGLDPLLTWQALGSETGVIEIGDEEAERLATASHHWREGIAGDFAAARTCL
ELHTPPDGEDLWELRFGLQAEADPSLKLPAAAAWAAGAEPLQLGEIRVDQPGEVLLEGMRALSVF
PAIERGLESATPETMQLTPAEAFVLVRTAARQLRDAGVGVELPPSLSGGLASRLGLSIKAELEPERS
SGFTLGECLAWEDLMIGGVTLTLRELERLSGKRSPLVRHKGAWIELRPNDLKNAERFCGAKPELS
LDDALRLTGTEGELLMRMPVHRFDAGPRLQSVLQQYHQQKAPDPLPAPEGFSGQLRPYQERGLGWL
AFLHRFDQGACLADDMGLGKTIQLLAFLLQHLKAEQELKRPVLLVAPTSVLTNWRREAESFTPELV
TEHYGPRRPSTPAELKKALKEVDLVLTSYGLLQRDSELLETQDWQGVVIDEAQAIKNPGAKQSQA
RDLARTGRIKSNRFRIALTGTPVENRVSELWALMDFLNPKVLGEEDFFRQRYRMPPIERYGDMSSLR
DLKGRVGPFILRLKTDKTIISDLPEKVELSEWVGLSKEQKSLYSKTVEDTLDAIARAPRGQRHGQ
VLALLTRLKQICNHPALALSEGAVDDGFLGRSAKLQRLEEILDEVIEAGDRALLFTQFAEWGHLR
AWMQQRWKSEVPFLHGGTRKNERQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWW
NPAVENQATDRAYRIGQTNRMVMHKFITSGSVEEKIDRMIREKSRLAEDVIGSGEDWLGLSLGGDQL
RDLVSLEDT

FIGURE 10 (continued)

85/96

SEQ ID NO: 89, *Synechococcus* sp. CC9902 Syn_sp_CC9902_SNF2 nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTTCCCGCCATTCGTACTTCCAGCAGTTCCGGACAGCCGGCA
CTGCTCATTTGGGCTGACACCTGGCGTGTGCGCTCACCGGAGGGGCCCCGGACTCACACCCGCTCTG
CATCCCTTCACCCTTGGCTCGGACGATCTCAAAGCTTGGTTGACCGAACGGGACCTGATGCCTGGG
GGCAGCATCGATGCCACCGCCTGCCTCACCTCCCAAGCCGCAGCGTCAAACCCCGCAAAAGTCGA
ACCCAACCGAGCGAACCAGCCCCAGAGGGACCGGCCTGGACCGGATTGCCAATGCAAGCAGGAGAG
CCCATTCCGAAGCAAATGGAATGGTGGCCCTGGCAGGTACAAGGCCTCGCGGTGGAGCCATCGGCC
GCAACGGAATGGCTCGCCCGTTTACCCCTATCGGGCCGACATCCAGACCTCGGAGATGAATTGCGC
TGGTGGAGCCATCTCCAACGTTGGTCCCTCAGCTTGGTGGCCCCGGGGGCGCTGGATTCCCCAGATG
GAATTAAGCAAAGGCGAGGGTTACCCCCACCGAGCGCGCTGGGTTCCTTGTGAACCGTGAGGAA
GATCGACGACGGCTCGAAGACCTCGCGGCCACGCTGCCCTCGTGGCGACCTGTGCCCTCCCTTGG
CGTGAGCCACTTGGACGCCGTAGCAACCGCACCAACAGGCTTCGACCGGAAGCGATGCGAGCCGCC
AACCCGGTGGCTTGCTGCCGCCCCCGGAGCGGTGCGCTCAGGGTGGCCACCTTGCTTGAAGACTTG
GTGGATGCAGAGCTGCGCAAGGGATTTGAACCCACCACAGAGGGGCTCGACCCCTACTCACCTTG
TGGCAAGAGGCCCTGGCCTCAGAAACCGGTGTTGTGGAGGTGGGCAACGAGGATGCAGAACGCCTT
ACCGCGGCAAGCCTGCACTGGCGCGAAGGGATTGCCGGAGGCTTCGCTGCTGCCCGCACCTGCCTC
GAACTAAACACCCCAAACGAAGGCGAAGAACTCTGGGACCTGAAGTTTGGCTTGCAAGCGGAGGCC
GATCCCAGCCTCAAGCTGCCGGCCGCGCGGCCTGGGCCTCAGGAGCCGAAACACTCCAGCTCGGG
GAGATCAAAGTTGACCAGGCGGGGGAAGTGCTGCTGGAGGGTCTTGGCCGAGCCCTCACGGTGTTC
CCTCCGATCGAACGCGGACTGGAAAGCGCAACGCCAGAAACGATGCAGCTCACGCCAGCGGAGGCG
TTTGTCTTGGTGCGAACAGCAACGCACCAAGCTCCGCAATGCCGGCATCGGCGTCGAACTGCCCCC
AGCCTTTCAGGGGGCCTCGCCAGCCGGCTTGGTTTAGCCATCAAGGCAGATTTACCAGATCGATCC
AGCGGCTTCACCCTCGGAGAATCTCTGGACTGGAGCTGGGATCTGATGATCGGCGGCGTCACACTC
ACCCTGCGAGAGCTCGAACGGCTCAGCGGTAAGCGCAGTCCGCTTGTGCGCCACAAGGGAGCCTGG
ATCGAACTGCGACCCAACGATCTCCGCAACGCCGAACGCTTCTGTGGAGCCAATCCAGAACTGAGC
CTCGACGATGCCCTAAGGCTCACGGCCACAGAAGGGGAGCTAATGATGCGCTTGCCGGTGCATCGC
TTTGATGCGGGGCTCGGCTTCAGGGAGTTCTCGAGCAATATCACCAGCAAAAAGCCCCCGATCCC
CTTCCCGCTCCAGAGGGATTTTCCGGACAACCTGCGTCCTTATCAAGAACGTGGCTTGGGCTGGCTG
GCCTTCTTACATCGCTTCGATCAAGGCGCCTGCCTGGCGGACGACATGGGCTTGGGCAAGACCATC
CAATTGTTGGCCTTCCTGCAGCACCTCAAAGCCGAGCACGAACTCAAACGCCCGGTGCTGTTGGTG
GCCCCAACCTCGGTGCTCACGAATTGGCGACGGGAGGCGGAAGCCTTCACCCCCGAGCTGTCGGTG
AAAGAGCACTACGGCCCACGCCGGCCTTCCACGCCGGCCGCTTGAAAAAAGAACTCAAAGATGTG
GATCTGGTGCTCACCAGTTACGGCCTGATGCAACGCGACAGCGAGCTGCTGGACAGCGTCGACTGG
CAAGGGGTGTGATCGACGAAGCGCAGGCGATCAAAAACCCTGGGGCGAAACAAAGCCAAGCAGCC
CGAGACCTGGCCCGAGCTGGAAAGAGCAGCAGGTTCCGCATCGCACTCACCGGCACACCGGTGGAA
AACCGCGTCAGCGAGCTGTGGGCGCTGATGGATTTCTCAACCCAAAGGTGTTGGGAGAGGAAGAC
TTCTTTCGTCAGCGCTACCGCATGCCAATTGAGCGCTACGGCGATATGTCGTCGTTACGCGATCTC
AAAGCGCGGGTCGGCCCCCTTCATCCTGCGCCGTCTCAAACCCGACAAGTCGATCATTTCCGACCTG
CCTGAAAAGGTGGAGCTCAGTGAATGGGTGGGTCTCAGCAAAGAACAGAAATCGCTGTACAACAAA
ACCGTTGAAGACACCCTCGACGCCATTGCCACCGCACCGCGGGGGCAACGCCATGGCCAGGTGCTA
GCCCTCTTGACCCGGTTAAAGCAGATTTGCAATCACCCGGCTTTAGCCCAACGCGAAGGGGGCCGTT
GACAGCGAATTCCTTGGCCGTTCCGCCAAGCTGATGCGACTCGAAGAAATCCTCGAAGAGGTGATT
GAAGCCGGCGATCGCGCTTTGCTATTACCCAATTGCGCGAATGGGGGCATCTCCTGCAGGCCTGG
ATGCAACAACGCTGGAAGTCTGAGGTTCCCTTCTGACGGCGGAACCCGCAAGAGTGATCGGCAA
GCGATGGTGGATCGATTCCAAGAGGACCCCCGGGGACCTCAACTCTTTCTTCTGTCCCTCAAGGCC
GGTGGTGTAGGCCTCAACCTCACCCGGGCCAGTCATGTGTTCCACGTGATCGCTGGTGAATCCA

FIGURE 10 (continued)

86/96

GCGGTGGAAAACCAAGCCACCGACCGGGCCTATCGAATTGGTCAAACCAACCGGGTAATGGTGCAC
AAATTCGTCACCCGTGGCTCGGTGGAAGAAAAAATCGACCAAATGATTCGTGAAAAAGCTCGAATG
GCTGAAGACGTGATCGGCTCCGGTGAAGACTGGCTCGGGAGCCTTGGCGGCGATCAGCTGCGCAAT
CTTGTGCCCCTCGAGGACACCTAA

SEQ ID NO: 90, *Synechococcus* sp. CC9902 Syn_sp_CC9902_SNF2 translated polypeptide

MSLLHATWLPARTSSSSGQPALLIWADTWRVASPEGPGLTPALHPFTLGSDDLKAWLTERDLMPG
GSIDATACLTLPSRSVKPRKSRTQPSEPAPEGPWTGLPMQAGEPIPKQMEWWPWQVQGLAVEPSA
ATEWLARLPLSGRHPDLGDELRWWSHLQRWSLSLVARGRWIPQMELSKGEGYPHRARWVPLLNREE
DRRRLDLAATLPLVATCALPWREPLGRRSNRTTRLRPEAMRAANPVACCRPRSGRLRVATLLEDL
VDAELRKGFEPTEGLDPLLLTLWQEALASETGVEVGNEDAERLTAASLHWREGIAGGFAAARTCL
ELNTPNEGEELWDLKFGLQAEADPSLKLPAAAAWASGAETLQLGEIKVDQAGEVLLEGLGRALT
PPIERGLSATPETMQLTPAEAFVLVRTATHQLRNAGIGVELPPSLSGGLASRLGLAIKADLPDRS
SGFTLGESLDWSWDLMIGGVTLTLRELERLSGKRSPLVRHKGAWIELRPNDLRNAERFCGANPELS
LDDALRLTATEGELMMRLPVHRFDAGPRLQGVLEQYHQKAPDPLPAPEGFSGQLRPYQERGLGWL
AFLHRFDQGACLADDMGLGKTIQLLAFLQHLKAEHELKRPVLLVAPTSVLTNWRREAEAFTELSV
KEHYGPRRPSTPAALKKELKDVDLVLTSYGLMQRDSELLDSVDWQGVVIDEAQAIKNPGAKQSQA
RDLARAGKSSRFRIALTGTPVENRVSELWALMDFLNPKVLGEEDFFRQRYRMPIERYGDMSSLRDL
KARVGPFILRLKTDKSIISDLPEKVELSEWVGLSKEQKSLYNKTVEDTLDAIATAPRGQRHGQVL
ALLTRLKQICNHPALAQREGAVDSEFLGRSAKLMRLEEILEEVIEAGDRALLFTQFAEWGHLQAW
MQQRWKSEVPFLHGGTRKSDRQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHVDRWWNP
AVENQATDRAYRIGQTNRMVHKFVTRGSVEEKIDQMIREKARMAEDVIGSGEDWLGSLGGDQLRN
LVALEDT

SEQ ID NO: 91, *Synechococcus* sp. RS9916 Syn_sp_RS9916_SNF2 nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTCCCGGCCATCCGTACACCCACCAGTTCCGGGCGTGCCGCC
CTGCTGGTGTGGGCGGACACCTGGCGTGTGGCGGAGCCGGCGGGCCCCGGCGTGACCCCGGCCACC
CATCCCTTCACCCTCAGCGCCGATGACCTGCGCGCCTGGCTGAGCGAACGGGAGCTGCTGCCCGAC
GGCATCATCGATGCCACCGCCTGCCTCACCTGCCAGCCGCACGGTGAAACCGAAGCGGAAGCGT
GGCGAGACCGCCCCTGTGGATGAGGGCTGGACGGGTCTGCCCCTGCAGGCGGGAGAACCGATTCCG
AAGCAGACCGAATGGTGGCCCTGGCAGGTACAGGGCCTGGCGGTCTGAACCCGGTGACGCCACCGCC
TGGCTGGCCCGCTTGCCCCCTCTCCGGCCGCCACCCCGACCTCGCCGATGAGCTGCGCTGGTGGAGC
CACATGCAGCGCTGGGCCCTCAGCCTGATTGCTCGCAGTCGCTGGATTCCCCAGGTGGAGCTGAGC
AAAGGGGAGGGCTACCCCCACCGCGCCCGTTGGGTGCCTCTGCTCAATCGCGAAGACGATCGCCGC
CGCCTGGAAGACATGGCGGCCCGCCTGCCGCTGGTGGCCACCTGCGCTCTCCCCTGGCGCGAACCC
ACCGGGAAGCGCAGCAACCGCACCCCGGCTGCGGCCTGAGGCGATGCGGGCGGCCAATCCGGTG
GCCTGTTGTCTGTCCTCCCGCAGCGGCCGACTGCGCGTCGCCACCCTGCTCGAAGACCTGGTGGATGCC
CAGCTGCGCACGGGTTTCACAGCCCAGACGGACGGGCTCGATCCCCTGCTTGCCGCCTGGGAGGAG
GCCCTCGGCAGCGACACCGGCGTGATCCACCTGGGCGATGAAGACGCAGAGCGTCTGGCCACCGCC
AGCCATCACTGGCGCGAAGGGGTGGCCGGCACTGTGGCGGCGGCGCGGGCCTGCCTGGAAGTGGAG
ACCCCGACGACGGCGATGACCTCTGGACCCTGCGGTTCGCACTGCAGGCCGAAGCGGATCCACG
CTCAAGGTGCCGGCCGCCCTCGCCTGGGCGGCCGGTCCGAAGGGACTCCAGCTCGGCGAAATCGCC
GTGGAGCATCCGGGCGAACTGCTGCTGGAAGGCATGGGCCGGGCGCTCACGGTGTTTCCACCGATC
GAACGCGGTCTCGACAGCGCCACGCCGAAGGGATGCAACTACCCCGCCGAAGCCTTCGTGCTG
GTGCGCACCGCAGCCCGCGAACTCCGCGATGTGGGGGTGGGCGTGGAGCTTCCAGCCAGCCTCTCG
GGTGGCCTGGCGAGCAGGCTCGGCCTGGCGATTGAGGCGGAAGTACCGGAGAAATCCCGCGGTTTC

FIGURE 10 (continued)

87/96

ACGCTGGGCGAAACCCTCGACTGGAGCTGGGAGCTGATGATCGGCGGCGTCAACCCTGACGCTGCGG
GAACTGGAGCGCCTGGCGGGCAAGCGCAGCCCCCTGGTGCGGCACAAGGGCACCTGGATCGAGCTG
CGCCCCAACGATCTCAAGAATGCGGAGCGGTTTTTCGCCGCGAAGCCCGATCTCAGCCTCGACGAT
GCCCTGCGCCTCACCGCCAGCGAAGGCGACACGCTGATGCGCATGCCGGTGACCGCCTGGAAGCG
GGCCACAGGCTGCAGGCGGTGCTCGAGCAGTATCACCAACAGAAAGCTCCCGATCCCCTGCCGGCG
CCGGAGGGCTTCTGCGGCCAGCTGCGGCCCTTACCAGGAGCGGGGCCCTCGGCTGGCTGGCCTTTCTG
CACCGCTTTGATCAAGGCGCCTGCCTGGCCGACGACATGGGTCTGGGCAAGACCATCCAGCTGCTC
GCCTTTCTGCAGCACCTGAAGGCCGAGCAGGAGCTGAAGAGGCCGGTGTTGCTCGTGGCGCCCACC
TCGGTGCTCACCAACTGGAAGCGGGAGGCCGCCGCTTCACGCCGGAGCTCGAGGTGAAGGAGCAC
TACGGGCCCAGGCGCCCTGCCACCCCTGCAGCACTCAAGAAGAGCCTCAAGGATGTGGATCTGGTG
CTCACCAGCTACGGCCTGCTCCAACGCGACAGCGAACTGCTCGAAAGTCTCGATTGGCAGGGGGTG
GTGATCGACGAAGCGCAGGCAATCAAGAATCCGAGCGCCAAACAGAGCATGGCGGCCCGAGACCTG
GCCCCGCGCAGGACGCAGCAGCCGTTTTCCGCATTGCCCTCACCGGCACGCCGGTGGAGAACCGGGTG
AGCGAGCTCTGGGCCTTGATGGATTTCTCAACCCGCGGGTGCTCGGCGAAGAGGACTTCTTCCGC
CAGCGCTACCGCATGCCGATTGAGCGCTATGGCGACATGTCGTCGCTGCGGGATCTGAAATCCCGC
GTGGGACCTTTTCATTCTTCGCCGGCTCAAAACCGACAAAGCGATCATTTCCGACCTGCCCCGAAAAG
GTGGAAGTGAAGCGAATGGGTGGGATTGAGCAGGGAGCAGAAAGCGCTCTATGCCAAAACCGTCGAG
GACACCCTCGATGCGATTGCCCCGGGCGCCCCGCGGACAACGGCATGGCCAGGTGCTGGGGTTGCTC
ACCAAGCTGAAGCAGATCTGTAACCATCCCGCCCTGGCCCTGAAAGAGGAGGCGGCCGCGACGAG
TTCCTGCAGCGCTCCATGAACTGCAGCGCCTGGAGGAAATCCTCGAGGAGGTGATCGACGCCGGC
GACCGCGCCCTGCTCTTCACCCAGTTCGCCGAATGGGGCCATCTGCTGCAGGGTTACCTGCAACGG
CGCTGGCGCAGCGAAGTGCCGTTCTTGAACGGCAGCACCAGCAAGAGCGAACGCCAGGCGATGGTC
GATCGCTTCCAGGAAGACCCGCGGGGGCCTCAGCTGTTCTGCTGTCACTGAAAGCCGGTGGTG
GGCCTCAACCTCACCCGCGCCAGCCATGTGTTTACATCGATCGCTGGTGGAATCCGGCGGTGGAA
AACCAGGCCACCGACCGCGCCTACCGGATCGGCCAGACGAACCGGGTGATGGTGACAAAGTTCATC
ACCAGTGGATCGGTGGAAGAAAAAATCGACCGGATGATCCGCGAGAAATCACGCCTCGCCGAAGAC
ATCATCGGCTCAGGCGAAGATTGGCTCGGCGGGCTCGACATGGGCCAGCTGAAGGAACTGGTGAGC
CTCGACGACAACGGATCACTTTTCAGCATGA

SEQ ID NO: 92, *Synechococcus* sp. RS9916 Syn_sp_RS9916_SNF2 translated polypeptide

MSLLHATWLPARTPTSSGRAALLVWADTWRVAEPAGPGVTPATHPFTLSADDLRAWLSERELLPD
GIIDATACLTLPSRTVKPKRKRGETAPVDEGWTGLPLQAGEPIPKQTEWWPWQVQGLAVEPGAATA
WLARLPLSGRHPDLADELRWWSHMQRWALSILIARSRWIPQVELSKGEGYPHRARWVPLLNREDDRR
RLEDMAARLPLVATCALPWREPTGKRSNRTRRLRPEAMRAANPVACCRPRSGRLRVATLLEDLVDA
QLRTGFTAQTDGLDPLLAWEELGSDTGVIHLGDEDAERLATASHHWREGVAGTVAAARACLELE
TPDDGDDLWTLRFALQAEADPTLKVPAALAWAAGPKGLQLGEIAVEHPGELLLEGMGRALTVPPI
ERGLDSATPEGMQLTPAEAFVLVRTAARELRDVGVGVELPASLSGGLASRLGLAIQAELEPKSRGF
TLGETLDWSWELMIGGVTLTLRELERLAGKRSPLVRHKGTWIELRPNDLKNAERFFAAKPDLSLDD
ALRLTASEGDTLMRMPVHRLEAGPRLQAVLEQYHQKAPDPLPAPEGFCGQLRPYQERGLGWLAFL
HRFDQGACLADDMGLGKTIQLLAFLQHLKAEQELKRPVLLVAPTSVLTNWKREAAFTPELEVKEH
YGPRRPATPAALKKSLKDVDLVLTSYGLLQRDSELLESLDWQGVVIDEAQAIKNPSAKQSMAARDL
ARAGRSSRFRIALTGTPVENRVSELWALMDFLNPRVLGEEDFFRQRYRMPPIERYGDMSSLRDLKSR
VGPFILRRLKTDKAIISDLPEKVELSEWVGLSREQKALYAKTVEDTLDAIARAPRGQRHGQVLGLL
TKLKQICNHPALALKEEAAGDEFLQSMKLQRLEEILEEVIDAGDRALLFTQFAEWGHLQGYLQR
RWRSEVPFLNGSTSKSERQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWNPAVE
NQATDRAYRIGQTNRMVMHKFITSGSVEEKIDRMIREKSRLAEDIIGSGEDWLGGLDMGQLKELVS
LDDNGSLSA

FIGURE 10 (continued)

88/96

SEQ ID NO: 93, *Synechococcus* sp. WH 7805 Syn_sp_WH7805_SNF2
nucleic acid sequence

ATGAGCCTGCTGCACGCCACCTGGCTACCCGCCATCCGCACTCCCAGCAGCTCCGGAAGGGCTGCT
TTGCTGGTATGGGCTGACACCTGGCGTGTGGCCGACCCCCCTCGGCCCCGGGGCCACACCCGCCCTT
CATCCGTTACCCCTGAGCGCGGAGGATCTGCGCGCCTGGCTCACAGAGCGCGATTTGCTTCCGGAC
GGAATCATCGATGCGACCGCATGCCTCACCCCTGCCGAGCCGCAGTGTGAAACCACGGCGGGCCCCGT
GGCTCAGCTGCCGCCACCCCCCTCATCAGAAGAGCAGCCCCCTTGGTGCGGGCTGCCGCTGCAAGCC
GGCGAACCGATCCCGAAAACACCGAGTGGTGGCCATGGCAGGTGCAGGGGCTGGCGATCGAACCG
ATGGCCGCCACGGCATGGCTGGCCAAGCTTCCACTGTCAGGCCATCACCCCTGATCTGGCCGATGAG
TTGCGCTGGTGGAGTCACATGCAGCGATGGGCCCTCAGTCTTGTGGCTAGGGGGCGCTGGCTGCCC
CAGGTGGAATTGAGCCGAGGTGAGGGGTATCCACACCGGGCCCCGCTGGGTCCCGCTTCTCAATCGA
GAGGAAGACCGGCGCCGCTGGAGGACCTTGCCGCCCGTCTGCCCCCTGGTTGCCACGTGTGCGTTG
CCCTGGAGAGAGCCACAGGAAAGCGCAGCAATCGCATCACCAAGGCTGCGCCCAGAGGCCATGCGC
GCTGCCAATCCCGTGGCCTGCTGTGCTCCCCGCAGCGGTTCGATTGCGGGTGGCCACATTGCTGGAG
GATCTGGTAGATGCCCAGCTGCGCAAGGGCTTCCATCCCGATGACGAGGGGCTCGACCCCTGCTC
TGCGCCTGGGAAAACGCCCTGAGTTCGGAGACCGGGGTGATCGATCTGAATGATGAAGATGCCGAA
CGCCTTGCCACGGCGAGCCACCACTGGCGCGAGGGAGTGGCTGGCAATGTGGCGGCTGCCAGGGCC
TGCCCTTGAACCTCGCCACACCGAACGAGGGGGAAGAGCTCTGGGATCTGCGCTTCTATCTGCAGGCC
GAAGCCGATCCAACGCTGAAGGTACCGGCCGGAGCAGCCTGGGCGCTGGACCCGAAGGCCCTTCAA
CTCGGGGAGATTCTGTGGAGCATCCCGGTGAGGTGCTGCTCGAAGGCATGGGGCGTGCTCTCACG
GTGTTTGAACCAATCGAACGGGGCCTGGATAGCGCCACGCCGGAAGCGATGCAGCTCACCCCGGCG
GAAGCCTTCGTGCTGGTGCACACCGCCGCCCGTCAGCTCCGGGACGTGGGCGTTGGTGTGGATCTC
CCTCCCAGCCTCTCGGGAGGCCTGGCCAGCCGCCTCGGTCTGGCGATCAAGGCCGAACCTACCCAAA
CGCTCGCGGGGGTTACCCCTTGGGGAAAATCTCGACTGGAACCTGGGAGCTGATGATCGGGGGCGTC
ACCCTGACGCTGCGGGAGCTGGAACGGCTGGCCGGCAAGCGCAGCCCCCTTGGTGCGCCACAAGGGG
GCCTGGATCGAACTCAGGCCCAATGATCTCAAAAATGCAGAACGATTCTGTGCCGCCAATCCTGAT
CTGAGCCTGGACGATGCCCTTCGCCTGACGGCCAGCGAAGGGGACACGCTGATGCGCCTCCCCGTT
CATGCCTTTGATGCTGGCCCTCGCCTTCAAGGGGTGTTGGAGCAATAACCACCAGCAGAAAGCACCG
GATCCACTTCCTGCGCCCGAGGGTTTCTGCGGTTCAGCTTCGCCCTTACCAGGAACGAGGCCTGGGC
TGGCTGGCCTTCCTGCACCGCTTCGATCAGGGAGCCTGCCTCGCCGACGACATGGGCCTGGGCAAG
ACGATCCAGCTGCTGGCCTTCCTCCAGCACCTGAAGATGGAACAAGAACTGAAACGGCCGGTGCTG
CTGGTGGCTCCACCTCCGTGCTCACCAACTGGAAACGGGAAGCCGCGGCCTTCACCCCCGAGCTC
ACAGTGCATGAGCACTACGGCCCCAAACGACCCTCCACCCCAGCAGCACTGAAAAAAGCCCTGAAA
GACGTTGACCTGGTGTCTACCAGCTACGGGCTTCTGCAAAGAGACAGTGAACCTGCTTGAAAGTTTC
GACTGGCAGGGAACCGTGATCGATGAAGCTCAGGCGATCAAGAACCCTTCGGCCAAGCAAAGCCAG
GCAGCCCGTGATCTGGCTCGCACCCGCAAGGGCTCCAGGTTCCGCATTGCCCTCACTGGCACACCG
GTTGAAAACAGAGTGAGCGAGCTCTGGGCCCTGATGGATTTCTCAATCCGAACGTGCTCGGCGAA
GAGGAATTTTCCGGCAGCGCTACCGCATGCCGATCGAACGCTATGGCGATATGTCGTCGCTTCGC
GATCTCAAGTCGCGGGTGGGACCATTCAATTCTGCGGCGCTTGAAAACCGACAAGGCGATCATCTCC
GACCTCCCCGAAAAAGTGGAGCTGAGTGAATGGGTGGGGCTGAGCAAGGAACAGAAGTCCCTTTAC
GCGAAAACCGTGGAGAACACCCTCGATGCCATCGCCCGAGCTCCCCGAGGCAAGCGTCACGGCCAG
GTGCTGGGACTGCTGACGCGCCTCAAACAGATCTGCAATCACCCGGCTCTGGCCTTAAAGGAAGAG
GTGGCAGGCGACGACTTCCTGCAGCGATCGGTGAAGCTGCAGCGGCTCGAAGAGATTCTCGAAGAG
GTGATTGCAGCGGGGGATCGAGCCCTGCTGTTACCCAGTTCGCGGAATGGGGGCATCTGCTGCAG
GGCTACCTGCAACGCCGCTGGCGCAGCGAGGTGCCGTTCTGAGCGGCAGCACTAGCAAAGGAGAA
CGTCAGGCCATGGTGGATCGCTTCCAGGAAGACCCGCGCGGGCCCCCAGCTGTTCTGTGTCCCTC
AAAGCCGGCGGTGTGGGATTGAACCTGACCCGGGCCAGCCACGTGTTCCACATCGACCGCTGGTGG

FIGURE 10 (continued)

89/96

AATCCTGCAGTTGAAAACCAGGCCACTGACCGTGCTTACCGGATTGGCCAGACCAATCGGGTGATG
GTGCATAAGTTCATCACCAGTGGCTCAGTGGAAGAGAAGATCGACCGGATGATCCGGGAGAAGTCC
AGACTGGCGGAAGACATCGTGGGCTCCGGCGAGGAGTGGCTCGGTGGCTTCGACATGGGCCAACTC
AAGGAGCTGGTGAGCCTCGAGGACAACGAAACACGCAACCCATGA

**SEQ ID NO: 94, *Synechococcus* sp. WH 7805 Syn_sp_WH7805_SNF2
translated polypeptide**

MSLLHATWLPARTPSSSGRAALLVWADTWVRVADPLGPGATPALHPFTLSAEDLRAWLTERDLLPD
GIIDATACLTLPSRSVKPRRPRGSAAATPSSEEQPPWCGLPLQAGEPIPKTTTEWWPWQVQGLAIEP
MAATAWLAKLPLSGHHPDLADELRWWSHMQRWALS LVARGRWLPQVELSRGEGYPHRARWVPLLNR
EEDRRRLEDLAARLPLVATCALPWREPTGKRSNRITRLRPEAMRAANPVACCRPRSGRLRVATLLE
DLVDAQLRKGFHPDDEGLDPLLCAWENALSSETGVIDLNDEDAERLATASHHWREGVAGNVAAARA
CLELATPNEGEELWDLRFYLQAEADPTLKVPAGAAWAAGPEGLQLGEIPVEHPGEVLLEGMGRALT
VFEPPIERGLDSATPEAMQLTPAEAFVLVRTAARQLRDVGVGVVDLPPSLSGGLASRLGLAIKAELPK
RSRGFTLGENLDWNWELMIGGVTLTLRELERLAGKRSPLVRHKGAWIELRPNDLKNAERFCAANPD
LSLDDALRLTASEGDTLMRLPVHAFDAGPRLQGVLEQYHQQKAPDPLPAPEGFCGQLRPYQERGLG
WLAFLHRFDQGACLADDMGLGKTIQLLAFLQHLKMEQELKRPVLLVAPTSVLTNWKREAAAFTEP
TVHEHYGPKRPSTPAALKKALKDVDLVLT SYGLLQRDSELLESFDWQGTVIDEAQA IKNPSAKQSQ
AARDLARTRKGSRFRIALTGTPVENRVSELWALMDFLNPNVLGEEFFRQRYRMP IERYGDMSSLR
DLKSRVGPFI LRRLKTDKAIISDLPEKVELSEWVGLSKEQKS LYAKTVENTLDAIARAPRGKRHGQ
VLGLLTRLKQICNHPALALKEEVAGDDFLQRSVKLQRLEEILEEVIAAGDRALLFTQFAEWGHL LQ
GYLQRRWRSEVPFLSGSTSKGERQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWW
NPAVENQATDRAYRIGQTNRMVMHKFITSGSVEEKIDRMIREKSRLAEDIVGSGEEWLGGFDMGQL
KELVSLEDNETRNP

**SEQ ID NO: 95, *Synechococcus* sp. WH 8102 Syn_sp_WH8102_SNF2
nucleic acid sequence**

ATGAGCCTGCTGCACGCCACCTGGCTTCCCGCCATCCGTACCTCTGGCAGTTCCGGCCAACCGGCA
CTGCTCATTTGGGCTGACACCTGGCGGGTGGCGACACCAGAGGGCCCCGGGCTAACTCCGGCGCTG
CACCCGTTACCCCTGGAACCCGACGACCTCAAGGCCTGGCTTCAGGAACGCGACCTGTTGCCAGGC
GGCAGCATCGATGCCACCGCCTGCCTCACCTGCCAGTCGCACGGTAAAACCCCGCAAGAGCCGC
AGCAAAACGGCCGAACCAGCGCCCGAAGAGCCCATCTGGACCGGTCTGCCGATGCAGGCCGGAGAG
CCGATTCCGAAACAGACAGAATGGTGGCCGTGGCAAGTCCAGGGCCTCGCTGTGCGAGCCCTCTGCC
GCCACGGAGTGGCTCTCACGCCTTCCCCTGTCAGGACGGAATCCAGACCTGGCCGATGAGCTGCGC
TGGTGGAGCCACCTGCAGCGCTGGGCCCTCAGCCTTGTGGCCCCGGGGGCGCTGGATTCCCCAGATG
GAACTGAGCAAAGGCGAGGGATATCCCCACCGGGGCCCGTTGGGTGCCTCTGCTCAACCGCGAGGAG
GACCGGCGACGTCTGGAGGATCTGGCCGCCAGCCTGCCGCTGGTGGCCACCTGCGCCCTGCCCTGG
CGGGAACCGATGGGTGCGCGCAGCAACCGCATGACACGGCTGCGTCCGGAGGCCATGCGTGCCGCC
AACCCGGTGGCCTGCTGCCGGCCCCGCGAGTGGCCGCCTGCGGGTGGCCACGCTGCTGGAGGATCTG
GTCGACGCACAGCTGCGCAAGGACTTTGAACCATCCACCGACGGCCTCGATCCCCTGTTGACCCTG
TGGCAAGACGCCCTGGGCTCCGAAACAGGGGTGATTGAGATCGGTGATGAACAGGCCGAACGGCTG
GCCAGCGCCAGCTTCCATTGGCGCGAGGGCATCGCTGGAGATTTGCGCCGCTGCACGCACCTGCCTG
GAACTGCAGACACCTGCAGAGGGAGAAGAGCTCTGGGAGCTGCGGTTTGGGCTGCAGGCCGAGTCG
GATCCGAGCCTCAAGCTGCCCCGCCGCTGCGGCCTGGGCCTCCGGTGCCGACCAACTCCAGTTGGGA
GAAGTGACAGTCGAGCAGCCCGGTGAAGTGCTGCTGGAGGGTCTGGGACGCGCCCTCACCGTGTTT
CCACCGATCGAAAGGGGCTGGAGACCGCTACGCCTGACACGATGCAGCTGACCCCCGCCGAAGCC
TTCGTGCTGGTGCGGACCGCAGCGCGGCAGCTGCGGGATGCCGGCGTCGGCGTCGACCTTCCCCC
AGCCTGTCGGGGGGCCTGGCCAGCCGCCTGGGTCTGGCGATCAAGGCGGAGCTGCCAGAGCGCTCC

FIGURE 10 (continued)

90/96

AGCGGCTTCAGCCTCGGCGAATCCCTCGACTGGAGCTGGGATCTGATGATCGGCGGGGTGACGCTC
ACCCTGCGGGAACTGGAGCGGTTGAGCGGCAAACGCAGCCCCCTCGTGCGCCACAAGGGGGCCTGG
ATCGAATTGCGACCGAACGATCTGAGAAACGCCGAACGCTTCTGCGGTGCCAACCCGGAGCTCAGC
CTGGACGATGCCCTGCGGATCACCGCCACCGAAGGCGATCTGCTGATGCGTCTGCCGGTGCATCGC
TTTGAGGCCGGCCCCAGGCTGCAGGCGGTGCTGGAGCAGTACCACCAGCAGAAGGCCCCGGATCCG
TTGCCAGCGCCGGAGGGGTTCTGCGGCCAGCTGCGGCCCTTACCAGGAGCGTGGCCTGGGCTGGCTG
GCCTTCCTCAACCGCTTCGACCAAGGCGCCTGCCTGGCGGACGACATGGGTCTGGGTAAGACCATC
CAGCTGCTGGCCTTCCTGCAGCACCTGAAAGCAGAGCAGGAAGTGAAGCGCCCGGTGCTGCTGGTG
GCCCCCACATCGGTGCTCACAACCTGGCGACGGGAAGCGGAAGCCTTCACCCCCGAAGTGGCGGTG
CGCGAGCACTACGGACCGCGGCGTCCCTCCACTCCGGCTGCGCTGAAGAAGGCGTTGAAGGATGTC
GACTTAGTCCTCACCAGCTACGGCCTACTGCAGAGGGACAGTGAATTGCTGGAGTCTCAGGATTGG
CAGGGGGTTGTGATCGATGAAGCCCAAGCGATCAAGAATCCCAGTGCCAAGCAGAGCCAGGCAGCC
CGAGACCTGGCCAGACCAGCCAAAGGCAACCGCTTCCGCATCGCCCTCACGGGCACACCGGTGGAG
AACAGGGTCAGCGAGCTCTGGGCTTTGATGGATTTCTCAGTCCCAAGGTGCTGGGAGAAGAAGAC
TTCTTCCGTCAGCGCTACCGGATGCCGATCGAGCGCTATGGCGACATGGCATCCCTACGGGACTTA
AAAGCCAGGGTCGGCCCCCTTCATCCTGCGCCGGCTGAAAACCGACAAGACGATCATTTCCGATCTG
CCCGAGAAGGTGGAACCTCAGCGAATGGGTGGGGTTGAGCAAGGAGCAGAAATCGCTGTACAGCAA
ACCGTTGAAGACACCCTGGATGCCATTGCCCGGGCGCCTCGTGGACAGCGCCATGGTCAGGTGCTG
GGACTGCTCACCCGCCTGAAGCAGATCTGCAACCATCCGGCCCTGGCATTGAGTGAAAACGCTGTT
GACGACGGCTTTCTGGGGCGCTCCGCCAAGTTGCAACGGCTTGAGGAAATCCTCGATGAGGTGATC
GAAGCAGGGGATCGGGCGCTGCTGTTACCCAGTTCGCCGAGTGGGGCCATCTGCTGCAGTCCTGG
ATGCAACAACGTTGGAAGGCGGATGTGCCCTTCTGCGATGGAGGGACGCGCAAAAACGAACGGCAG
GCCATGGTGGATCGTTTTTCAGGAGGACCCCCGCGGCCCGCAGCTGTTCTGCTGTCGCTCAAAGCC
GGCGGGGTGGGTCTGAACCTGACCAGGGCCAGCCACGTGTTCCACATCGATCGCTGGTGGAAACCCT
GCGGTAGAGAACCAGGCCACCGACCGTGCTTATCGGATCGGCCAGACCAACCGGGTGATGGTGCAC
AAATTCATCACAAGCGGATCCGTAGAAGAAAAAATTGACCGGATGATCCGAGAGAAGTCGCGCCTG
GCAGAGGATGTGATCGGTTCCGGTGAAGACTGGCTCGGGTGCCTGGCCGGTGATCAGCTGCGCAAT
CTCGTTGCCCTGGAGGACACCTGA

SEQ ID NO: 96, *Synechococcus* sp. WH 8102 yn_sp_WH8102_SNF2 translated polypeptide

MSLLHATWLPARTSGSSGQPALLIWADTWRVATPEGPGLTPALHPFTLEPDDLKAWLQERDLLPG
GSIDATACLTLPSRTVKPRKSRSKTAEPAPEEPIWTGLPMQAGEPIPKQTEWWPWQVQGLAVEPSA
ATEWLSRLPLSGRNPDLADELRWWSHLQRWALS LVARGRWIPQMELSKGEGYPHRARWVPLLNREE
DRRRLDLAASLPLVATCALPWREPMGRRSNRMTRLRPEAMRAANPVACCRPRSGRLRVATLLEDL
VDAQLRKDFEPSTDGLDPLLT LWQDALGSETGVIEIGDEQAERLASASFHWREGIAGDFAAARTCL
ELQTPAEGEELWELRFGLQAESDPSLKLPAAAAWASGADQLQLGEVTVEQPGEVLLEGLGRALTVF
PPIERGLETATPDTMQLTPAEAFVLVRTAARQLRDAGVGVDLPSSLSSGLASRLGLAIKAELPERS
SGFSLGESLDWSWDLMIGGVTLTLRELERLSGKRSPLVRHKGAWIELRPNDLRNAERFCGANPELS
LDDALRITATEGDLLMRLPVHRFEAGPRLQAVLEQYHQKAPDPLPAPEGFCGQLRPYQERGLGWL
AFLNRFDQGACLADDMGLGKTIQLLAFLQHLKAEQELKRPVLLVAPTSVLTNWRREAEFTPELAV
REHYGPRRPSTPAALKKALKDVDLVLT SYGLLQRDSELLESQDWQGVVIDEAQAIAKNPSAKQSQA
RDLARPAKGNRFRIALTGT PVENRVSELWALMDFLSPKVLGEEDFFRQRYRMPIERYGDMASLRDL
KARVGPFILRLKTDKTIISDLPEKVELSEWVGLSKEQKSLYSKTVEDTLDAIARAPRGQRHGQVL
GLLTRLKQICNHPALALSENAVDDGFLGRSAKLQRLEEILDEVIEAGDRALLFTQFAEWGHL LQSW
MQQRWKADV PFLHGGTRKNERQAMVDRFQEDPRGPQLFLLSLKAGGVGLNLTRASHVFHIDRWNP
AVENQATDRAYRIGQTNRVMVHKFITSGSVEEKIDRMIREKSRLAEDVIGSGEDWLGCLAGDQLRN
LVALED T

FIGURE 10 (continued)

91/96

**SEQ ID NO: 97, *Synechococcus elongatus* PCC 6301 Synel_PCC6301_SNF2
nucleic acid sequence**

ATGGCAGTGCTGCACGGTGGCTGGCTCGGCGATCGCTTCTGCGTTTGGGCCGAGGCTTGGCAGGCT
GGTGAGCCTCAGTCGGCAGCAGAAATTGCGATTTCATCCCTACGCGATCGCGGCCACTGACTTAAAT
GATTGGTGCCAGAAGTACCGTCTGGGATCCCTGACGGGGACGCCAACAGAAGTCCTGCTCTCTATT
CCCAGTGACCTGAAGAAAGAGGCGGTTCTACCGTTTCTGAGTGGTCAGGAAATTCCAGATGGGGCG
CTGCTTTTGGTCTTGGCAGATCCCCGTGCTGTCGCTAGAAGCCGCGATCGCCGGTCAATGGCTGGCG
ACCTTGCCGCTGGGTTCGGCGGAGGATCATCCTTGGCTGGGGCCAGATCTACGCTTTTGGAGCCAC
ATCTACCGCTGGGCACAAAGTTTGGCTGGCTCGGGGGCGCTTTTATCCGGCGCTGGAGTCGAGCGAT
CGCGGTTTAAACGGCAGTTTGGTTGCCACTGTTTAATCAAGCGGGCGATCGCCAGCGCTTCGATCGC
TATAGTCAGCAGCTGCCCTTTAGTCAGTTTTGCTATCAGGCAATCGAAACAGCGGCAGCTTGTCCT
TGGCAGCCTCAACCGCAGGATCTGTTGCTGCGAGTCCTACAGACTTGGTTGACAGCACGACTACAA
CCGGCGATCGCGGCGGGAACCTCTCGTGTCTGCTGATCTGCTGGCGGCTTGGCAGCAATCGCTAGCG
AATGGAAAACCGCTAAAGCTAGAAGACAGTGAAGCCAGTCGCTTGCAAACGGCGATCGATCGCTGG
TTACTACCAGTGCAAGATGGCGCAGCTCAGGCTTGGCGGATGGTTTTTGCGCCCTTGTCGCCCTACG
GAGCAAGAGCAGCCCTGGCAATTGGAGTTTGGCTTACAAGCAGCGACCGATCCCGATCGCTTTCGG
CCGGCCTCTCTCCTCTGGCAGGATCCGCTGCCACCTGGGCTACCAGATCAATCTCAGGAATTGCTG
TTACGCGGCTTGGGACAGGCTTGTGCGGCTCTATCCCCAATTGCAAACCAGTCTGGCGACAGCCTGT
CCAGAATTCCATCCACTGACCACAGCGGAGGTCTATCAGCTGCTCAAGCAGGTGATTCCTCAGTGG
CAAGAGCAGGGCATTGAAGTGCAACTGCCGCCGGGCTTGCCTGGTCAAGGGCGACACCGGCTGGGA
GTGGAAGTCAGCGCCACGTTGCCGAGCGATCGCCCGAGTGTGGGGCTGGAAGCACTACTGCAGTTT
CGTTGGGAGCTGAGTCTGGGCGGTGAGCGGCTGACCAAAGCAGAAGTGGAACGCTTGGCAGCCCTG
GAAACGCCCTTGGTGGAAATCAACGGCGACTGGATTGAGGTGCGGCCGAGGATATTGAGTCGGCG
CGAGAGTTTTTCCGTAAGCGCAAGGATCAGCCAAATTTGACCTTGGCGGATGCGATCGCGATCGCC
AGTGGTGAGTCGCCGAATGTTGGTGCCTGCCGGTGGTCAATTTTGAAGCGGCGGGCTTACTCGAA
GAAGCCTTGGCCGTGTTTCAGGGGCAGCGATCGCCTGCGGCTTTGCCCGCTCCGCCACCTTTTCAG
GGCGAGCTGCGACCCCTATCAAGAGCGGGGGGTGGGCTGGCTCAGCTTTTTTGCAGCGCTTCGGGATT
GGGGCTTGCCCTGCCGACGACATGGGCTTGGGTAAGACGATTGAGCTGCTGGCCTTTTTTACTGCAT
CTCAAACACAGCAACGAGCTGACGCGGCCGGTGCTGCTAGTCTGTCCGACTTCGGTGCTGGGCAAC
TGGGAACGGGAGGTGCAGAAATTTGCACCGGAGCTTCGCTGGAAGCTGCACTATGGCCCCGATCGC
GCTCAGGGTAAGGCTTTGGCGACAGCGCTCAAGGACTGCGATTTGGTGCTGACCAGTTACTCCTTG
GTGGCGCGAGATCAGAAAGCGATCGCGGCGATCGACTGGCAAGGCATTGTGCTGGATGAAGCCCAG
AACATCAAGAATGACCAGGCGAAACAGACGCAGGCGGTGCGAGCGATCGCCCAAAGTCCGACGCAA
AAGCCCCGCTTTCGGATTGCCCTGACAGGGACGCCGGTTGAGAATCGCCTCAGTGAGTTGTGGTCG
ATTGTCGAGTTTTTTGCAGCCGGGACATTTAGGCACCAAGCCATTCTTTCAAAGCGCTTTGTACG
CCGATCGAGCGTTTTTGGCGATGCGGATTTCGCTGACAGCATTGCGGCAGCGCGTGCAACCGTTAATC
CTACGGCGACTGAAAACCGATCGCAGCATTATTGCCGACTTGCCCTGAGAAGCAAGAAATGACGGTC
TTTTGTCCGTTGGTACAGGAGCAGGCCGATCGCTATCAGGTGCTAGTCAATGAAGCGCTAGCCAAT
ATTGAAGCAAGTGAAGGCATTCAGCGGCGCGGCCAGATTTTGGCATTGCTAACGCGACTGAAGCAG
CTCTGTAATCATCCGTCGTTGTTGCTCGAAAAGCCGAAGCTCGATCCGAATTTTGGCGATCGCTCA
GCCAAGTTGCAGCGCTTACTAGAAATGTTGGCGGAGCTAACGGATGCGGGCGATCGCGCTTTGGTG
TTTACGCAGTTTGCGGGCTGGGGTAGTTTTGCTGCAGCAATTTTTTGCAGGAACAGCTAGGGCGAGAG
GTGCTGTTTTTTGTCGGGCAGTACCAAGAAGGGCGATCGCCAACAGATGGTTGATCGCTTCCAAAAT
GATCCGCAGGCACCGGCAATTTTTCATCCTGTCATTGAAGGCTGGCGGGGTGGGGCTCAACCTGACG
AAAGCCAATCATGTCTTTCATTACGATCGCTGGTGGAATCCGGCAGTTGAAAACCAAGCGACCGAT
CGCGCGTTTTCGGATTGGGCAACGACGCAATGTACAGGTGCACAAGTTTGTCTGCGCTGGCACTCTA
GAAGAAAAAATTGATCAGATGATCGCTAGCAAGCAAGCATTAGCACAGCAGATTGTGCGGTAGTGGT
GAGGATTGGCTAACGGAACTAGACACCAATCAACTCCGGCAACTCTTGATCCTCGATCGCTCAGCT
TGGGTAGAAGAGGAAGAGCCTTAG

FIGURE 10 (continued)

92/96

SEQ ID NO: 98, *Synechococcus elongatus* PCC 6301 Synel_PCC6301_SNF2 translated polypeptide

MAVLHGGWLGDRFCVWAEAWQAGEPQSAAEIAIHPYAIAATDLNDWCQKYRLGSLTGTPTEVLLSI
PSDLKKEAVLPFLSGQEIPDGALLWSWQIPVLSLEAAIAGQWLATLPLGSAEDHPWLGPDLRFWSH
IYRWAQSLLARGRFYPALESSDRGLTAVWLPLFNQAGDRQRFDRYSQQLPFSQFCYQAIETAAACP
WQPQPQDLLLRVLQTWLTARLQPAIAAGTLVSADLLAAWQQSLANGKPLKLEDSEASRLQTAIDRW
LLPVQNGAAQAWRMVLRLLVPPTQEQQPWQLEFGLQAATDPDRFRPASLLWQDPLPPGLPDQSQELL
LRGLGQACRLYPQLQTSLATACPEFHPLTTAEVYQLLKQVIPQWQEQGIEVQLPPGLRGQGRHRLG
VEVSATLPSDRPSVGLLEALLQFRWELSLGGQRLTKAEVERLAALETPLVEINGDWIEVRPQDIESA
REFFRKRKDQPNLTLADAIAIASGESPNVGRLPVVNFEAAGLLEEALAVFQGQORSPAALPAPPTFQ
GELRPYQERGVGWSLFLQRFGIGACLADDMGLGKTIQLLAFLLHLKHSNELTRPVLLVCPTSVLGN
WEREVQKFAPELRWKLHYGPDRAQGKALATALKDCDLVLTSSVSLVARDQKAIAAIDWQGIVLDEAQ
NIKNDQAKQTQAVRAIAQSPTQKPRFRIALTGTPVENRLSELWSIVEFLQPGHLGTPFFQKRFT
PIERFGDADSLTALRQRVQPLILRRLKTDRSIIADLPEKQEMTVFCPLVQEQADRYQVLVNEALAN
IEASEGIQRRGQILALLTRLKQLCNHPSLLLEKPKLDPNFGDRSAKLQRLLEMLAELTDAGDRALV
FTQFAGWGSLLQQFLQEQLGREVLFLSGSTKKGDRQQMVDRFQNDPQAPAIIFILSLKAGGVGLNLT
KANHVHFHYDRWWNPAVENQATDRAFRIGQRRNVQVHKFVCAGTLEEKIDQMIASKQALAQQIVGSG
EDWLTELDTNQLRQLLILDRSAWVEEEEP

SEQ ID NO: 99, *Synechococcus elongatus* PCC 7942 Synel_PCC7942_SNF2 nucleic acid sequence

ATGGCAGTGCTGCACGGTGGCTGGCTCGGCGATCGCTTCTGCGTTTGGGCCGAGGCTTGGCAGGCT
GGTGAGCCTCAGTCGGCAGCAGAAATTGCGATTCATCCCTACGCGATCGCGGCCACTGACTTAAAT
GATTGGTGCCAGAAGTACCGTCTGGGATCCCTGACGGGGACGCCAACAGAAGTCCTGCTCTCTATT
CCCAGTGACCTGAAGAAAGAGGCGGTTCTACCGTTTCTGAGTGGTCAGGAAATTCCAGATGGGGCG
CTGCTTTGGTCTTGGCAGATCCCCGTGCTGCTACTAGAACCCGCGATCGCCGGTCAATGGCTGGCG
ACCTTGCCGCTGGGTTTCGGCGGAGGATCATCCTTGGCTGGGGCCAGATCTACGCTTTTGGAGCCAC
ATCTACCGCTGGGCACAAAGTTTGGTGGCTCGGGGGCGCTTTTATCCGGCGCTGGAGTCGAGCGAT
CGCGGTTTAAACGGCAGTTTGGTTGCCACTGTTTAATCAAGCGGGCGATCGCCAGCGCTTCGATCGC
TATAGTCAGCAGCTGCCCTTTAGTCAGTTTTGCTATCAGGCAATCGAAACAGCGGCAGCTTGTCCT
TGGCAGCCTCAACCGCAGGATCTGTTGCTGCGAGTCCTACAGACTTGGTTGACAGCACGACTACAA
CCGGCGATCGCGGCGGGAACTCTCGTGTCTGCTGATCTGCTGGCGGCTTGGCAGCAATCGCTAGCG
AATGGAAAACCGCTAAAGCTAGAAGACAGTGAAGCCAGTCGCTTGCAAACGGCGATCGATCGCTGG
TTACTACCAGTGCAGAATGGCGCAGCTCAGGCTTGGCGGATGGTTTTTGCGCCTTGTCCTCCGCTACG
GAGCAAGAGCAGCCCTGGCAATTGGAGTTTGGCTTACAAGCAGCGACCGATCCCGATCGCTTTTGG
CCGGCCTCTCTCCTCTGGCAGGATCCGCTGCCACCTGGGCTACCAGATCAATCTCAGGAATTGCTG
TTACGCGGCTTGGGACAGGCTTGTGCGGCTCTATCCCCAATTGCAAACCAGTCTGGCGACAGCCTGT
CCAGAATTCCATCCACTGACCACAGCGGAGGTCTATCAGCTGCTCAAGCAGGTGATTCCTCAGTGG
CAAGAGCAGGGCATTGAAGTGCAACTGCCGCCGGGCTTGGCGTGGTCAAGGGCGACACCGGCTGGGA
GTGGAAGTCAGCGCCACGTTGCCGAGCGATCGCCCGAGTGTGGGGCTGGAAGCACTACTGCAGTTT
CGTTGGGAGCTGAGTCTGGGCGGTGACGCGGCTGACCAAAGCAGAAGTGGAACGCTTGGCAGCCCTG
GAAACGCCCTTGGTGGAAATCAACGGCGACTGGATTGAGGTGCGGCCGAGGATATTGAGTCGGCG
CGAGAGTTTTTCCGTAAGCGCAAGGATCAGCCAAATTTGACCTTGGCGGATGCGATCGCGATCGCC
AGTGGTGAGTCGCCGAATGTTGGTGCCTGCCGGTGGTCAATTTTGAAGCGGCGGGCTTACTCGAA
GAAGCCTTGGCCGTGTTTCAGGGGCAGCGATCGCCTGCGGCTTTGCCCGCTCCGCCACCTTTTCAG
GGCGAGCTGCGACCCTATCAAGAGCGGGGGGTGGGCTGGCTCAGCTTTTTGCAGCGCTTCGGGATT
GGGGCTTGCTCGCCGACGACATGGGCTTGGGTAAGACGATTGAGCTGCTGGCCTTTTTTACTGCAT
CTCAAACACAGCAACGAGCTGACGCGGCCGGTGCTGCTAGTCTGTCCGACTTCGGTGCTGGGCAAC

FIGURE 10 (continued)

93/96

TGGGAACGGGAGGTGCAGAAATTTGCACCGGAGCTTCGCTGGAAGCTGCACTATGGCCCCGATCGC
GCTCAGGGTAAGGCTTTGGCGACAGCGCTCAAGGACTGCGATTTGGTGCTGACCAGTTACTCCTTG
GTGGCGCGAGATCAGAAAGCGATCGCGGCGATCGACTGGCAAGGCATTGTGCTGGATGAAGCCCAG
AACATCAAGAATGACCAGGCGAAACAGACGCAGGCGGTGCGAGCGATCGCCCAAAGTCCGACGCAA
AAGCCCCGCTTTTCGGATTGCCCTGACAGGGACGCCGGTTGAGAATCGCCTCAGTGAGTTGTGGTCTG
ATTGTCTGAGTTTTTTGCAGCCGGGACATTTAGGCACCAAGCCATTCTTTCAAAGCGCTTTGTCTACG
CCGATCGAGCGTTTTTGGCGATGCGGATTTCGCTGACAGCATTGCGGCAGCGCGTGCAACCGTTAATC
CTACGGCGACTGAAAACCGATCGCAGCATTATTGCCGACTTGCCCTGAGAAGCAAGAAATGACGGTC
TTTTGTCCGTTGGTACAGGAGCAGGCCGATCGCTATCAGGTGCTAGTCAATGAAGCGCTAGCCAAT
ATTGAAGCAAGTGAAGGCATTCAGCGGCGCGGCCAGATTTTGGCATTGCTAACGCGACTGAAGCAG
CTCTGTAATCATCCGTCGTTGTTGCTCGAAAAGCCGAAGCTCGATCCGAATTTTGGCGATCGCTCA
GCCAAGTTGCAGCGCTTACTAGAAATGTTGGCGGAGCTAACGGATGCGGGCGATCGCGCTTTGGTG
TTTACGCAGTTTTCGGGGCTGGGGTAGTTTGTCTGCAGCAATTTTTTGCAGGAACAGCTAGGGCGAGAG
GTGCTGTTTTTGTCTGGGCGAGTACCAAGAAGGGCGATCGCCAACAGATGGTTGATCGCTTCCAAAAT
GATCCGCAGGCACCGGCAATTTTTCATCCTGTCATTGAAGGCTGGCGGGGTGGGGCTCAACCTGACG
AAAGCCAATCATGTCTTTCATTACGATCGCTGGTGGAAATCCGGCAGTTGAAAACCAAGCGACCGAT
CGCGCGTTTTTCGGATTGGGCAACGACGCAATGTACAGGTGCACAAGTTTGTCTGCGCTGGCACTCTA
GAAGAAAAAATTGATCAGATGATCGCTAGCAAGCAAGCATTAGCACAGCAGATTGTCTGGTAGTGGT
GAGGATTGGCTAACGGAAGTAGACACCAATCAACTCCGGCAACTCTTGATCCTCGATCGCTCAGCT
TGGGTAGAAGAGGAAGAGCCTTAG

**SEQ ID NO: 100, *Synechococcus elongatus* PCC 7942 Synel PCC7942
SNF2 translated polypeptide**

MAVLHGGWLGDRFCVWAEAWQAGEPQSAAEIAIHPYAIAATDLNDWCQKYRLGSLTGTPTEVLLSI
PSDLKKEAVLPFLSGQEIPDGALLWSWQIPVLSLEAAIAGQWLATLPLGSAEDHPWLGPDLRFWSH
IYRWAQSLLARGRFYPALESSDRGLTAVWLPLFNQAGDRQRFDRYSQQLPFSQFCYQAIETAAACP
WQPQPQDLLLLRVLQTLTLARLQPAIAAGTLVSADLLAAWQQSLANGKPLKLEDSEASRLQTAIDRW
LLPVQNGAAQAWRMVLRLLVPPTEQEQPWQLEFGLQAATDPDRFWPASLLWQDPLPPGLPDQSQELL
LRGLGQACRLYPQLQTSLATACPEFHPLTTAEVYQLLKQVIPQWQEQGIEVQLPPGLRGQGRHRLG
VEVSATLPSDRPSVGLLEALLQFRWELSLGGQRLTKAEVERLAALETPLVEINGDWIEVRPQDIESA
REFFRKRKQPNLTLADAIAIASGESPNVGRLPVVNFEAAGLLEEALAVFQGQSPALPAPPTFQ
GELRPYQERGVGWSLFLQRFQIGACLADDMGLGKTIQLLAFLHLKHSNELTRPVLLVCPTSVLGN
WEREVQKFAPELRWKLHYGPDRAQGKALATALKDCDLVLTSSVSLVARDQKAIAAIDWQGIIVLDEAQ
NIKNDQAKQTQAVRAIAQSPTQKPRFRIALTGTPVENRLSELWSIVEFLQPGHLGTPFFQKRFT
PIERFGDADSLTALRQRVQPLILRLKTDRSIIADLPEKQEMTVFCPLVQEQADRYQVLVNEALAN
IEASEGIQRRGQILALLTRLKQLCNHPSLLLEKPKLDPNFGDRSAKLQRLLEMLAELTDAGDRALV
FTQFAGWGSLLQQFLQEQLGREVLFLSGSTKKGDRQQMVDRFQNDPQAPAIIFILSLKAGGVGLNLT
KANHVVFHYDRWWNPAVENQATDRAFRIGQRRNVQVHKFVCAGTLEEKIDQMIASKQALAQQIVGSG
EDWLTELDTNQLRQLLILDRSAWVEEEEP

**SEQ ID NO: 101, *Thermosynechococcus elongatus* BP-1 Theel_BP-1_SNF2
nucleic acid sequence**

ATGGCTATTTTCCATGGCACATGGCTCCCAGAGCCGGCGCCACAGTTTTTTCATTTGGGCGGAAGAA
TGGCGATCGCTGGCTCAGGCAATCACGCCTTGGGCTCCCCCGGCGATTCCGGTTTATCCCTACGCC
ACCCAGAGAAAAACACCTCTTAGGAAGACAGCCCGCCCAAGTGCCACCTACGTTGCTTTACCGGCC
CAGATTCAGGGGCATCAACTGTTACCACCACCGCTGGCGGAAGTGCAGGGGGAACTCCTATTTTGT
TGGCAGGTGCCCCGGCTGGTCAATTCCCGCTTCAGAAAGTTTTAGAACAACTGCATCAACTGAGTCTT
CACGGCCAAGACAGTGGCAGTATTGGCGATGATTTGCGCTATTGGCTGCACGTGAGTCGCTGGTTG

FIGURE 10 (continued)

94/96

CTGGATTTAATTGTGCGTGGCCAATACCTGCCAACACCAGAGGGCTGGCGGATTCTGCTGACCCAC
GGGGGCGATCGCGATCGCCTGCGCCACTTCAGCCAATTGATGCCGGATCTGTGTCGCTGTTATCAA
GCCGATGGCACAGCGTTGCAGTTGCCACCCCATGCTGCAGATCTCCTGGCGGATTTTCTACAGCAC
ACCCTACAGGGTTATCTCCACACTGCCCTTGCTGACCTCGAATTGCCCAAAGTAGGCTTAGCCAAA
GAACATGGCCACTGGCTAGCCTTCCTGAAAACGGGTCAAACCCCGGAACTGCCACCTCCCCTCAT
GAACGCCTGCACCGCTGGCAAGAACCCTACCGCGAGCAGTTGCATCTGCGTCCCCAATGGCGACTG
GCTCTGCAATTGGTTCCCCCAGATACTGCCGATGGTGACTGGCACTTGGCCTTTGGGCTGCAAACG
GAAGGGGAAACGGACACCATGCTAAGGGCCGCCGAGATTTGGCAATGCACCCAAGAGGCCCTCCTC
TATCAAGGGCAGGTGCTCTGGCAGCCCCAAGAAACCCTGTTGCGGGGACTGGGCTTGGCCTCCCGC
ATCTATCGTCCCCTCGATCGCAGTCTTCAAGAACGCTCCCCCGTGGCTCTGACTTTGCACACCACG
GAAGTTTATGCCTTCTTGCAAAGTGCAATTGCGCCCCCTTGAGCAGCAGGGGGTTGCGATCATTTTG
CCACCGAGTCTGCGCCGCAATAGCGCCCAACATCGCTTGGGTCTGAAAATAATTGCCACATTGCCG
CCGCCGGCCACTAACGGCTTGACGATTGACAGCTTGATGCAGTTTCAGTGGCAGTTGCAGTTGGGG
CAGCATCCCCTCTCGGAGGCGGATTTTGATCAACTGCGCCGCCAAGGGACGCCCTGGTTTATCTC
AATGGTGAGTGGGTCTTGCTGCGCCCCCAAGAGGTCAAGGCCGCTCAAGAGTTTCTCCAGTCTCCC
CCAAAGACCCAACTCTCCCTTGACAGAGACACTGCGCATTGCTACGGGGGATACGGTAACGGTGGCC
AAGTTGCCGATTCTTGCTTAGACACCAATGATGCACTCCAGACCCTCTTGATGGCCTCACGGGC
AAACAAAGCCTTGATCCAGTGCCAACACCGCAGGAGTTTTGCGGTGAACTGCGCCCCCTACCAGGCA
CGGGGGGTGGCGTGGCTGAGTTTCTTGGAACGCTGGCGGCTGGGGGCTTGCTTGGCGGACGATATG
GGCTTGGGGAAAACCATTTCAACTGTTGGCCTTTTTTGCTCCACCTCAAGGAAACGGGACGGGCCTAC
CGACCGACACTGTTGATCTGTCCTACCTCGGTGCTGGGGAACTGGCTGCGGGAGTGCCAAAAGTTT
GCCCCAACCTTGCGGGCCTATGTCCACCATGGGAGCGATCGCCCCAAGGGCAAGGCATTTCTGAAA
AAGGTTGAAACTCACGATCTAATTTTGACCAGTTATGCCCTCCTCCAGCGCGATCGCACACCTTG
CAGCAGGTTCTGTGGCAGCATTTGGTACTGGATGAAGCCCCAAAACATCAAGAATGCCAACACCCAG
CAGTCCCAAGCAGCGCGGGAACCTTTCGCCCCAGTTTCGCATTGCCCTGACGGGAACCCCCCTAGAA
AACCGCCTCCTCGAACTTTGGTCCATTATGGACTTCCTCCATCCGGGGTACTTGGGCCATCGCACC
TACTTTCAACACCGCTATGTCCGTCCCATTGAACGCTATGGCGACACCACCTCCCTCAATGCTCTG
CGCACCTATGTCCAGCCCTTTATTCTGCGGCGCCTGAAAACCGACCGCAGTATTATTCAAGACCTG
CCGGAAAAACAGGAGATGCTGGTGTATTGTGGCCTCACCTAGAGCAGATGCAGCTTTACACTGCT
GTGGTGGAAGACTCCCTTGCTGCTATCGAAAATAGTCAAGGCATTCAGCGGCGGGGCAATATCTTG
GCCACCCTGACCAAGTTGAAGCAAATCTGTAACCATCCCGCCCAGTATCTCAAGCAAGAAGACTAT
GCCCCCGATCGCTCAGGTAAATTGCAACGGCTTATAGAAATGCTGCAAGCGCTTCAGGAAGTGGGC
GATCGCGCCCTTGCTCTTTACCCAATTTGCCGAGTTTGGCACCCACCTGAAAACCTATCTGGAAAAG
GCGCTCCAGCAGGAGGTGTTTTCTCTCAGGACGCACCCCCAAAGCCCAGCGGGAACATCATGGTG
GAACGCTTTCAACACGATCCCGAGGCCCCCCAGGGTCTTTATTCTTTCCCTCAAGGCAGGGGGCGTC
GGTCTCAATTTGACTCGCGCTAACCATGTCTTTCACTACGATCGCTGGTGGAACCCAGCGGTAGAA
AATCAGGCCAGCGATCGCGTCTTCCGCATTGGTCAGGCCCCGCAATGTCCAAATCCATAAATTTATC
TGCACGGGTACCCTCGAAGAAAAGATCCACGAGCAAATCGAACAGAAAAAGCCCTTGCGGAAATG
ATTGTGGGTAGTGGCGAACACTGGCTGACTGAACTCAACCTCGACCAGTTGCGGCAACTGCTCACC
TTAGACAAAGAGCGGCTGATCACCTCTAG

SEQ ID NO: 102, *Thermosynechococcus elongatus* BP-1 Theel_BP-1_SNF2 translated polypeptide

MAIFHGTWLPEPAPQFFIWAEWRSLAQAITPWAPPAIPVYPYATQRKTPLRK TARPSATYVALPA
QIQGHQLLPPLAEVQGELLFLWQVPGWSIPASEVLEQLHQLSLHGQDSGSIGDDLRYWLHVSRWL
LDLIVRGQYLPTPEGWRILLTHGGDRDRLRHFSQLMPDLCRCYQADGTALQLPPHAADLLADFLQH
TLQGYLHTALADLELPKVGLAKEHGHWLAFLKTGQTPELPPPLIERLHRWQEPYREQLHLRPQWRL
ALQLVPPDTADGDWHLAFLQTEGETDTMLRAAEIWQCTQEALLYQGQVLWQPQETLLRGLGLASR

FIGURE 10 (continued)

95/96

IYRPLDRSLQERSPVALTLHTTEVYAFQLQSAIAPLEQQGVAILPPSLRRNSAQHRLGLKIIATLP
PPATNGLTIDSLMQFQWQLQLGQHPLSEADFDQLRRQGTPLVYLNGEWVLLRPQEVKAAQEFLQSP
PKTQLSLAETLRIATGDTVTVAKLPILGLDTNDALQTLLDGLTGKQSLDPVPTPQEFCEGELRPYQA
RGVAWLSFLERWRLGACLADDMGLGKTIQLLAFLHLKLTGRAYRPTLLICPTSVLGNWLRECQKF
APTLRAYVHHGSDRPGKGAFLKKVETHDLILTSYALLQRDRTTLQQVLWQHLVLDEAQNIKNANTQ
QSQAARELSAQFRIALTGTPLENRLLELWSIMDFLHPGYLGHRITYFQHRYVRPIERYGDTTSLNAL
RTYVQPFILRRLKTDRSIIQDLPEKQEMLVYCGLTLEQMQLYTAVVEDSLAAIENSQGIQRRGNIL
ATLTKLKQICNHQAQYLKQEDYAPDRSGKLQRLIEMLQALQEVGDRAVFTQFAEFGTHLKTYLEK
ALQQEVFFLSGRTPKAQRELMVERFQHDPEAPRVFILSLKAGGVGLNLTRANHVHFHYDRWWNPAAVE
NQASDRVFRIGQARNVQIHKFICTGTLEEKIHEQIEQKKALAEIVGSGEHWLTELNLQDLRQLLT
LDKERLITL

SEQ ID NO: 103, Motif 1

LADDMGLGK (T/S)

SEQ ID NO: 104, Motif 1a

L (L/V/I) (V/I/L) (A/C) P (T/M/V) S (V/I/L) (V/I/L) XNW

SEQ ID NO: 105, Motif 2

DEAQ (N/A/H) (V/I/L) KN

SEQ ID NO: 106, Motif 3

A (L/M) TGTPXEN

SEQ ID NO: 107, Motif 4

(L/I) XF (T/S) Q (F/Y)

SEQ ID NO: 108, Motif 5

S (L/V) KAGG (V/T/L) G (L/I) (N/T) LTXA (N/S/T) HV

SEQ ID NO: 109, Motif 5a

DRWWNPAAVE

SEQ ID NO: 110, Motif 6

QA (T/S) DR (A/T/V) (F/Y) R (I/L) GQ

SEQ ID NO: 111, ATPase domain of SEQ ID NO: 2

LADDMGLGKTPQLLAFLHLAAEDMLVKPVLIVCPTSVLSNWGHEINKFAPQLKTLLHHGDRRKKG
QPLVKQVKDQQIVLTSYALLQRDFSSLLKLVWQGVLDLDEAQNIKNPQAKQSQAARQLPAGFRIALT
GTPVENRLTELWSILEFLNPGFLGNQSFFQRRFANPIEKFGDRQSLILRLNLVRPFILRRLKTDQT
IIQDLPEKQEMTVFCDSLSEQAGLYQQLVESLQAIADSEGIQRHGLVLTLLTKLKQVCNHPDLLL
KKPAITHGHQSGKLIRLAEMLEEIISEGDRVLIFTQFASWGHLKPYLEKYFNQEVLYLHGGTPAE
QRQALVERFQQDPNSPYLFILSLKAGGTGLNLTRANHVHFVDRWWNPAAVENQATDRAFRIGQTRNV
QVHKFVCTGTLEEKINAMMADKQQALAEQTVDAENWLTRLDTDKLRQLLTLSATPVDYQAEASD

FIGURE 10 (continued)

96/96

SEQ ID NO: 112, *Oryza sativa* beta-expansin promoter

AAAACCACCGAGGGACCTGATCTGCACCGGTTTTGATAGTTGAGGGACCCGTTGTGTCTGGTTTTTC
CGATCGAGGGACGAAAATCGGATTCGGTGTAAGTTAAGGGACCTCAGATGAACTTATTCGGGAGC
ATGATTGGGAAGGGAGGACATAAGGCCCATGTCGCATGTGTTTGGACGGTCCAGATCTCCAGATCA
CTCAGCAGGATCGGCCGCGTTCGCGTAGCACCCGCGGTTTGATTTCGGCTTCCCGCAAGGCGGCGGC
CGGTGGCCGTGCCGCCGTAGCTTCCGCCGGAAGCGAGCACGCCGCCGCCGACCCGGCTCTGCG
TTTGCACCGCCTTGCACGCGATACATCGGGATAGATAGCTACTACTCTCTCCGTTTTCACAATGTAA
ATCATTCTACTATTTTCCACATTCATATTGATGTTAATGAATATAGACATATATATCTATTTAGAT
TCATTAACATCAATATGAATGTAGGAAATGCTAGAATGACTTACATTGTGAATTGTGAAATGGACG
AAGTACCTACGATGGATGGATGCAGGATCATGAAAGAATTAATGCAAGATCGTATCTGCCGCATGC
AAAATCTTACTAATTGCGCTGCATATATGCATGACAGCCTGCATGCGGGCGTGTAAGCGTGTTTCAT
CCATTAGGAAGTAACCTTGTCATTACTTATAACCAGTACTACATACTATATAGTATTGATTTTCATGA
GCAAATCTACAAAACCTGGAAAGCAATAAGAAATACGGGACTGGAAAAGACTCAACATTAATCACCA
AATATTTTCGCCTTCTCCAGCAGAATATATATCTCTCCATCTTGATCACTGTACACACTGACAGTGT
ACGCATAAACGCAGCAGCCAGCTTAAGTGTCTCACCCTCGCACACTGGCCTTCCATCTCAGGC
TAGCTTTCTCAGCCACCCATCGTACATGTCAACTCGGCGCGCGCACAGGCACAAATTACGTACAAA
ACGCATGACCAAATCAAAACCAACCGGAGAAGAATCGCTCCCGCGCGCGGGCGGCGACGCGCACGTAC
GAACGCACGCACGCACGCCCAACCCACGACACGATCGCGCGCGACGCCGGCGACACCGGCCGTCC
ACCCGCGCCCTCACCTCGCCGACTATAAATACGTAGGCATCTGCTTGATCTTGTCATCCATCTCAC
CACCAAAAAAAAAAAGGAAAAAAAAAACAAAACACACCAAGCCAAATAAAAGCGACAA

SEQ ID NO: 113, Prm 08774

GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACAATGGCGACTATCCACGGTAATTGG

SEQ ID NO: 114, Prm 08779

GGGGACCACTTTGTACAAGAAAGCTGGGTTC AATCGGACGCTTCGGCTT

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 September 2008 (04.09.2008)

PCT

(10) International Publication Number
WO 2008/104598 A3

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)

(74) Agent: MISTRY, Meeta; Basf Se, Global Intellectual Property, Gvx - C006, 67056 Ludwigshafen (DE).

(21) International Application Number:
PCT/EP2008/052450

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:
28 February 2008 (28.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
07103271.8 28 February 2007 (28.02.2007) EP
07104197.4 15 March 2007 (15.03.2007) EP
60/896,050 21 March 2007 (21.03.2007) US
60/909,510 2 April 2007 (02.04.2007) US

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): CROPDESIGN N.V. [BE/BE]; Technolopark 3, B-9052 Zwijnaarde (BE). CROP FUNCTIONAL GENOMICS CENTER [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu, Seoul 151-921 (KR).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): CHOI, Yang Do [KR/KR]; Banppo Tower Apt 102-301, Jamwon-dong, Seocho-ku, Seoul 137-909 (KR). PARK, Youn-II [KR/KR]; Jijok dong 919, 705-401, Youseong, Daejeon 305-330 (KR). CHUNG, Suck Won [KR/KR]; Noeun 559-10, Youseong, Daejeon 305-325 (KR). HWANG, In Gyu [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu, Seoul 151-921 (KR). OH, Jonghee [KR/KR]; San 56-1, Shillim 9-Dong, Gwanak-Gu, Seoul 151-921 (KR).

(88) Date of publication of the international search report:
13 November 2008

(54) Title: PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND A METHOD FOR MAKING THE SAME

(57) Abstract: The present invention relates generally to the field of molecular biology and concerns a method for enhancing various economically important yield-related traits in plants. More specifically, the present invention concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a Harpin-associated Factor G polypeptide (hereinafter termed HpaG"). The present invention also concerns plants having modulated expression of a nucleic acid encoding an HpaG polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs comprising HpaG-encoding nucleic acids, useful in performing the methods of the invention. The present invention also provides a method for enhancing yield-related traits in plants relative to control plants, by modulating (preferably increasing) expression in a plant of a nucleic acid sequence encoding a SWITCH 2/ SUCROSE NON-FERMENTING 2 (SWI2/SNF2) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs useful in performing the methods of the invention.

WO 2008/104598 A3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/052450

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBASE, CHEM ABS Data, EMBL, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	REN HAIYING ET AL: "Combinative effects of a bacterial type-III effector and a biocontrol bacterium on rice growth and disease resistance" JOURNAL OF BIOSCIENCES (BANGALORE), vol. 31, no. 5, December 2006 (2006-12), pages 617-627, XP002445065 ISSN: 0250-5991 the whole document	1-11,15, 17,22
X	DATABASE WPI Week 200159 Thomson Scientific, London, GB; AN 2001-530414 XP002445791 & CN 1 300 547 A (UNIV NANJING AGRIC) 27 June 2001 (2001-06-27) abstract	1-11,15, 17,22

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

29 May 2008

Date of mailing of the international search report

18/09/2008

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Bilang, Jürg

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2008/052450

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE WPI Week 200415 Thomson Scientific, London, GB; AN 2004-152686 XP002445085 -& KR 2003 068 302 A (CHOI J W) 21 August 2003 (2003-08-21) cited in the application abstract</p>	1-11,15, 17,22
A	<p>-----</p> <p>DATABASE WPI Week 200652 Thomson Scientific, London, GB; AN 2004-169939 XP002445082 & CN 1 225 559 C (UNIV NANJING AGRIC) 2 November 2005 (2005-11-02) abstract</p>	1-11,15, 17,22
A	<p>-----</p> <p>DATABASE WPI Week 200649 Thomson Scientific, London, GB; AN 2004-169928 XP002445083 & CN 1 219 059 C (UNIV NANJING AGRIC) 14 September 2005 (2005-09-14) abstract</p>	1-11,15, 17,22
A	<p>-----</p> <p>PENG JIAN-LING ET AL: "Expression of harpinXoo in transgenic tobacco induces pathogen defense in the absence of hypersensitive cell death" PHYTOPATHOLOGY, vol. 94, no. 10, October 2004 (2004-10), pages 1048-1055, XP002445066 ISSN: 0031-949X figure 1</p>	
A	<p>-----</p> <p>KIM JUNG-GUN ET AL: "Mutational analysis of Xanthomonas harpin HpaG identifies a key functional region that elicits the hypersensitive response in nonhost plants" JOURNAL OF BACTERIOLOGY, vol. 186, no. 18, September 2004 (2004-09), pages 6239-6247, XP002445067 ISSN: 0021-9193 page 6242</p> <p>-----</p> <p style="text-align: center;">-/--</p>	

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/052450

C(Continuation): DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>LIU FENGQUAN ET AL: "The internal glycine-rich motif and cysteine suppress several effects of the HpaG(Xooc) protein in plants"</p> <p>PHYTOPATHOLOGY, vol. 96, no. 10, October 2006 (2006-10), pages 1052-1059, XP008081958 ISSN: 0031-949X page 1053 page 1056, right-hand column - page 1057, right-hand column</p> <p>-----</p>	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2008/052450

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers allsearchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Invention 1: claims 1-11, 15, 17, 22, all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-11, 15, 17, 22, all partially

a method for enhancing yield related traits comprising modulating the expression of a nucleic acid encoding a HpaG protein, wherein said HpaG protein is represented by the sequence shown in SEQ ID NO: 2

Inventions 2 to 12: claims 1-11, 15, 17, 22, all partially

a method for enhancing yield related traits comprising modulating the expression of a nucleic acid encoding a HpaG protein, wherein in each separate invention said HpaG protein is represented by one of the sequences shown in table A, i. e. SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 28

Invention 13: claims 12-22, all partially

a plant comprising a nucleic acid encoding a HpaG protein, wherein said HpaG protein is represented by the sequence shown SEQ ID NO: 2 and the corresponding constructs

Inventions 14 to 24: claims 12-22, all partially

a plant comprising a nucleic acid encoding a HpaG protein, wherein in each separate invention said HpaG protein is represented by one of the sequences shown in table A, i. e. SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 28, and the corresponding constructs

Invention 25: claims 23-47

A method for enhancing yield-related traits comprising increasing the expression of in a plant of a nucleic acid sequence encoding a SWI2/SNF2 polypeptide, the corresponding plants and constructs

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2008/052450

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
CN 1300547	A	27-06-2001	NONE	
KR 2003068302	A		NONE	
CN 1225559	C	02-11-2005	CN 1451750 A	29-10-2003
CN 1219059	C	14-09-2005	CN 1451664 A	29-10-2003